Proceedings of the

International Congress of Mathematicians

Rio de Janeiro 2018

VOLUME II

Invited Lectures

Boyan Sirakov
Paulo Ney de Souza
Marcelo Viana
Editors

SBM

World Scientific

Proceedings of the

International Congress of Mathematicians

Rio de Janeiro 2018

Proceedings of the International Congress of Mathematicians
August 1 – 9, 2018, Rio de Janeiro, Brazil

Printed in Singapore

Published and Distributed by

World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

Tel: 65-6466-5775
Fax: 65-6467-7667
www.wspc.com
sales@wspc.com

ISBN: 978-981-3272-91-0 (volume print)
ISBN: 978-981-3272-87-3 (set print)
ISBN: 978-981-3272-88-0 (set ebook)

Proceedings of the **International Congress of Mathematicians**

Rio de Janeiro 2018

VOLUME II

Invited Lectures

Boyan Sirakov
Paulo Ney de Souza
Marcelo Viana
Editors

SBM

World Scientific

Contents

1 Logic and Foundations

2 Algebra

3 Number Theory

4 Algebraic and Complex Geometry

5 Geometry

6 Topology

8 Analysis and Operator Algebras

13 Combinatorics

14 Mathematical Aspects of Computer Science

15 Numerical Analysis and Scientific Computing

17 Mathematics in Science and Technology

18 Mathematics Education and Popularization of Mathematics

19 History of Mathematics

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (19–42)

ON NUMBERS, GERMS, AND TRANSSERIES

Matthias Aschenbrenner, Lou van den Dries and Joris van der Hoeven

Abstract

Germs of real-valued functions, surreal numbers, and transseries are three ways to enrich the real continuum by infinitesimal and infinite quantities. Each of these comes with naturally interacting notions of *ordering* and *derivative*. The category of H-fields provides a common framework for the relevant algebraic structures. We give an exposition of our results on the model theory of H-fields, and we report on recent progress in unifying germs, surreal numbers, and transseries from the point of view of asymptotic differential algebra.

Contemporaneous with Cantor's work in the 1870s but less well-known, P. du Bois-Reymond [1871, 1872, 1873, 1875, 1877, 1882] had original ideas concerning non-Cantorian infinitely large and small quantities (see Ehrlich [2006]). He developed a "calculus of infinities" to deal with the growth rates of functions of one real variable, representing their "potential infinity" by an "actual infinite" quantity. The reciprocal of a function tending to infinity is one which tends to zero, hence represents an "actual infinitesimal".

These ideas were unwelcome to Cantor (see Fisher [1981]) and misunderstood by him, but were made rigorous by F. Hausdorff [1906a,b, 1909] and G. H. Hardy [1910, 1912a,b, 1913]. Hausdorff firmly grounded du Bois-Reymond's "orders of infinity" in Cantor's set-theoretic universe (see Felgner [2002]), while Hardy focused on their differential aspects and introduced the *logarithmico-exponential functions* (short: *LE-functions*). This led to the concept of a *Hardy field* (Bourbaki [1951]), developed further mainly by Rosenlicht [1983a,b, 1984, 1987, 1995] and Boshernitzan [1981, 1982, 1986, 1987]. For the role of Hardy fields in *o-minimality* see Miller [2012].

Surreal numbers were discovered (or created?) in the 1970s by J. H. Conway [1976] and popularized by M. Gardner, and by D. E. Knuth [1974] who coined the term "surreal number". The surreal numbers form a proper class containing all reals as well as Cantor's ordinals, and come equipped with a natural ordering and arithmetic operations turning them into an ordered field. Thus with ω the first infinite ordinal, $\omega - \pi$, $1/\omega$, $\sqrt{\omega}$ make

The first-named author was partially support by NSF Grant DMS-1700439.
MSC2010: primary 03C64; secondary 12J15, 12J20, 34M15.

sense as surreal numbers. In contrast to non-standard real numbers, their construction is completely canonical, naturally generalizing both Dedekind cuts and von Neumann's construction of the ordinals. (In the words of their creator Conway [1994, p. 102], the surreals are "the only correct extension of the notion of real number to the infinitely large and the infinitesimally small.") The surreal universe is very rich, yet shares many properties with the real world. For example, the ordered field of surreals is real closed and hence, by Tarski [1951], an elementary extension of its ordered subfield of real numbers. (In fact, every set-sized real closed field embeds into the field of surreal numbers.) M. Kruskal anticipated the use of surreal numbers in asymptotics, and based on his ideas Gonshor [1986] extended the exponential function on the reals to one on the surreals, with the same first-order logical properties; see van den Dries and Ehrlich [2001a,b]. Rudiments of analysis for functions on the surreal numbers have also been developed by Alling [1987], Costin, Ehrlich, and Friedman [2015], and Rubinstein-Salzedo and Swaminathan [2014].

Transseries generalize LE-functions in a similar way that surreals generalize reals and ordinals. Transseries have a precursor in the *generalized power series* of Levi-Civita [1892–93, 1898] and Hahn [1907], but were only systematically considered in the 1980s, independently by Écalle [1992] and Dahn and Göring [1987]. Écalle introduced transseries as formal counterparts to his "analyzable functions", which were central to his work on Dulac's Problem (related to Hilbert's 16th Problem on polynomial vector fields). Dahn and Göring were motivated by Tarski's Problem on the model theory of the real field with exponentiation. Transseries have since been used in various parts of mathematics and physics; their formal nature also makes them suitable for calculations in computer algebra systems. Key examples of transseries are the *logarithmic-exponential series* (*LE-series* for short), see van den Dries, Macintyre, and Marker [1997, 2001]; more general notions of transseries have been introduced by van der Hoeven [1997] and Schmeling [2001]. A transseries can represent a function of a real variable using exponential and logarithmic terms, going beyond the more prevalent asymptotic expansions in terms of powers of the independent variable. Transseries can be manipulated algebraically—added, subtracted, multiplied, divided—and like power series, can be differentiated term-wise: they comprise a differential field. However, they carry much more structure: for example, by virtue of its construction, the field of LE-series comes with an exponential function; there is a natural notion of composition for transseries; and differential-compositional equations in transseries are sometimes amenable to functional-analytic techniques van der Hoeven [2001].

The logical properties of the *exponential* field of LE-series have been well-understood since the 1990s: by Wilkie [1996] and van den Dries, Macintyre, and Marker [1997] it is model-complete and o-minimal. In our book Aschenbrenner, van den Dries, and van der Hoeven [2017a] we focused instead on the *differential* field of LE-series, denoted below by $\mathbb{T}$, and obtained some decisive results about its model theory. Following A. Robinson's general ideas we placed $\mathbb{T}$ into a suitable category of *H-fields* and, by developing the extension theory of H-fields, showed that $\mathbb{T}$ is existentially closed as an H-field: each system of algebraic differential equations and inequalities over $\mathbb{T}$ which has a solution in an H-field extension of $\mathbb{T}$ already has one in $\mathbb{T}$ itself. In Aschenbrenner, van den Dries, and van der Hoeven [ibid.] we also prove the related fact that $\mathbb{T}$ is model-complete; indeed, we obtain a quantifier elimination (in a natural language) for $\mathbb{T}$. As a consequence, the elementary theory of $\mathbb{T}$ is decidable, and model-theoretically "tame" in various ways: for example, it has Shelah's *non-independence property* (NIP).

Results from Aschenbrenner, van den Dries, and van der Hoeven [ibid.] about existential closedness, model completeness, and quantifier elimination substantiate the intuition, expressed already in Écalle [1992], that $\mathbb{T}$ plays the role of a *universal domain* for the part of asymptotic differential algebra that steers clear of oscillations. How far does this intuition lead us? Hardy's field of LE-functions embeds into $\mathbb{T}$, as an ordered differential field, but this fails for other Hardy fields. The natural question here is: *Are all maximal Hardy fields elementarily equivalent to $\mathbb{T}$?* It would mean that any maximal Hardy field instantiates Hardy's vision of a maximally inclusive and well-behaved algebra of oscillation-free real functions. Related is the issue of embedding Hardy fields into more general differential fields of transseries. Positive answers to these questions would tighten the link between germs of functions (living in Hardy fields) and their transseries expansions. We may also ask how surreal numbers fit into the picture: *Is there a natural isomorphism between the field of surreal numbers and some field of generalized transseries?* This would make it possible to differentiate and compose surreal numbers as if they were functions, and confirm Kruskal's premonition of a connection between surreals and the asymptotics of functions.

We believe that answers to these questions are within grasp due to advances in our understanding during the last decade as represented in our book Aschenbrenner, van den Dries, and van der Hoeven [2017a]. We discuss these questions with more details in Sections 3, 4, 5. In Section 1 we set the stage by describing Hardy fields and transseries as two competing approaches to the asymptotic behavior of non-oscillatory real-valued functions. (Section 5 includes a brief synopsis of the remarkable surreal number system.) In Section 2 we define H-fields and state the main results of Aschenbrenner, van den Dries, and van der Hoeven [ibid.].

We let m, n range over $\mathbb{N} = \{0, 1, 2, \dots\}$. Given an (additive) abelian group A we let $A^{\neq} := A \setminus \{0\}$. In some places below we assume familiarity with very basic model theory, for example, on the level of Aschenbrenner, van den Dries, and van der Hoeven [ibid., Appendix B]. "Definable" will mean "definable with parameters".

1 Orders of Infinity and Transseries

Germs of continuous functions. Consider continuous real-valued functions whose domain is a subset of $\mathbb{R}$ containing an interval $(a, +\infty)$, $a \in \mathbb{R}$. Two such functions *have the same germ* (at $+\infty$) if they agree on an interval $(a, +\infty)$, $a \in \mathbb{R}$, contained in both their domains; this defines an equivalence relation on the set of such functions, whose equivalence classes are called *germs*. Addition and multiplication of germs is defined pointwise, giving rise to a commutative ring $\mathcal{C}$. For a germ g of such a function we also let g denote that function if the resulting ambiguity is harmless. With this convention, given a property P of real numbers and $g \in \mathcal{C}$ we say that $P\big(g(t)\big)$ holds *eventually* if $P\big(g(t)\big)$ holds for all sufficiently large real t in the domain of g. We identify each real number r with the germ of the constant function $\mathbb{R} \to \mathbb{R}$ with value r. This makes the field $\mathbb{R}$ into a subring of $\mathcal{C}$.

Following Hardy we define for $f, g \in \mathcal{C}$,

$$f \preccurlyeq g \quad :\Longleftrightarrow \quad \text{for some } c \in \mathbb{R}^{>0} \text{ we have } |f(t)| \leqslant c|g(t)| \text{ eventually,}$$
$$f \prec g \quad :\Longleftrightarrow \quad \text{for every } c \in \mathbb{R}^{>0} \text{ we have } |f(t)| < c|g(t)| \text{ eventually.}$$

The reflexive and transitive relation $\preccurlyeq$ yields an equivalence relation $\asymp$ on $\mathcal{C}$ by setting $f \asymp g :\Longleftrightarrow f \preccurlyeq g$ and $g \preccurlyeq f$, and $\preccurlyeq$ induces a partial ordering on the set of equivalence classes of $\asymp$; these equivalence classes are essentially du Bois-Reymond's "orders of infinity". Thus with x the germ of the identity function on $\mathbb{R}$:

$$0 \prec 1 \prec \log\log x \prec \log x \prec \sqrt{x} \prec x \asymp -2x + x\sin x \prec x^2 \prec \mathrm{e}^x.$$

One way to create interesting subrings of $\mathcal{C}$ is via expansions of the field of real numbers: any such expansion $\tilde{\mathbb{R}}$ gives rise to the subring $H(\tilde{\mathbb{R}})$ of $\mathcal{C}$ consisting of the germs of the continuous functions $\mathbb{R} \to \mathbb{R}$ that are definable in $\tilde{\mathbb{R}}$.

Hausdorff fields. A *Hausdorff field* is by definition a subfield of $\mathcal{C}$. Simple examples are

(1) $$\mathbb{Q}, \quad \mathbb{R}, \quad \mathbb{R}(x), \quad \mathbb{R}(\sqrt{x}), \quad \mathbb{R}(x, \mathrm{e}^x, \log x).$$

That $\mathbb{R}(x, \mathrm{e}^x, \log x)$ is a Hausdorff field, for instance, follows from two easy facts: first, an element f of $\mathcal{C}$ is a unit iff $f(t) \neq 0$ eventually (and then either $f(t) > 0$ eventually or $f(t) < 0$ eventually), and if $f \neq 0$ is an element of the subring $\mathbb{R}[x, \mathrm{e}^x, \log x]$ of $\mathcal{C}$, then $f \asymp x^k\, \mathrm{e}^{lx}(\log x)^m$ for some $k, l, m \in \mathbb{N}$. Alternatively, one can use the fact that an expansion $\tilde{\mathbb{R}}$ of the field of reals is o-minimal iff $H(\tilde{\mathbb{R}})$ is a Hausdorff field, and note that the examples above are subfields of $H(\mathbb{R}_{\exp})$ where $\mathbb{R}_{\exp}$ is the exponential field of real numbers, which is well-known to be o-minimal by Wilkie [1996].

Let H be a Hausdorff field. Then H becomes an ordered field with (total) ordering given by: $f > 0$ iff $f(t) > 0$ eventually. Moreover, the set of orders of infinity in H is totally ordered by $\preccurlyeq$: for $f, g \in H$ we have $f \preccurlyeq g$ or $g \preccurlyeq f$. In his landmark paper, Hausdorff [1909] essentially proved that H has a unique algebraic Hausdorff field extension that is real closed. (Writing before Artin and Schreier [1927], of course he doesn't use this terminology.) He was particularly interested in "maximal" objects and their order type. By Hausdorff's Maximality Principle (a form of Zorn's Lemma) every Hausdorff field is contained in one that is maximal with respect to inclusion. By the above, maximal Hausdorff fields are real closed. Hausdorff also observed that maximal Hausdorff fields have uncountable cofinality; indeed, he proved the stronger result that the underlying ordered set of a maximal Hausdorff field H is η_1: *if A, B are countable subsets of H and $A < B$, then $A < h < B$ for some $h \in H$.* A real closed ordered field is $\aleph_1$-saturated iff its underlying ordered set is η_1. Standard facts from model theory (or Erdős, Gillman, and Henriksen [1955]) now yield an observation that could have been made by Hausdorff himself in the wake of Artin and Schreier [1927]:

Corollary 1.1. *Assuming* CH (*the Continuum Hypothesis*), *all maximal Hausdorff fields are isomorphic.*

This observation was in fact made by Ehrlich [2012] in the more specific form that under CH any maximal Hausdorff field is isomorphic to the field of surreal numbers of countable length; see Section 5 below for basic facts on surreals. We don't know whether here the assumption of CH can be omitted. (By Esterle [1977], the negation of CH implies the existence of non-isomorphic real closed η_1-fields of size $2^{\aleph_0}$.) It may also be worth mentioning that the intersection of all maximal Hausdorff fields is quite small: it is just the field of real algebraic numbers.

Hardy fields. A Hardy field is a Hausdorff field whose germs can be differentiated. This leads to a much richer theory. To define Hardy fields formally we introduce the subring

$$\mathcal{C}^n := \{f \in \mathcal{C} : f \text{ is eventually } n \text{ times continuously differentiable}\}$$

of $\mathcal{C}$, with $\mathcal{C}^0 = \mathcal{C}$. Then each $f \in \mathcal{C}^{n+1}$ has derivative $f' \in \mathcal{C}^n$. A *Hardy field* is a subfield of $\mathcal{C}^1$ that is closed under $f \mapsto f'$; Hardy fields are thus not only ordered fields but also differential fields. The Hausdorff fields listed in (1) are all Hardy fields; moreover, for each o-minimal expansion $\tilde{\mathbb{R}}$ of the field of reals, $H(\tilde{\mathbb{R}})$ is a Hardy field. As with Hausdorff fields, each Hardy field is contained in a maximal one. For an element f of a Hardy field we have either $f' > 0$, or $f' = 0$, or $f' < 0$, so f is either eventually strictly increasing, or eventually constant, or eventually strictly decreasing. (This may fail for f in a Hausdorff field.) Each element of a Hardy field is contained in the intersection

$\bigcap_n \mathcal{C}^n$, but not necessarily in its subring $\mathcal{C}^\infty$ consisting of those germs which are eventually infinitely differentiable. In a Hardy field H, the ordering and derivation interact in a pleasant way: if $f \in H$ and $f > \mathbb{R}$, then $f' > 0$. Asymptotic relations in H can be differentiated and integrated: for $0 \neq f, g \not\asymp 1$ in H, we have $f \preccurlyeq g$ iff $f' \preccurlyeq g'$.

Extending Hardy fields. Early work on Hardy fields focused on solving algebraic equations and simple first order differential equations: Borel [1899], Hardy [1912a,b], Bourbaki [1951], Marić [1972], Sjödin [1971], Robinson [1972], Rosenlicht [1983a]. As a consequence, every Hardy field H has a smallest real closed Hardy field extension $\operatorname{Li}(H) \supseteq \mathbb{R}$ that is also closed under integration and exponentiation; call $\operatorname{Li}(H)$ the *Hardy-Liouville closure* of H. (Hardy's field of LE-functions mentioned earlier is contained in $\operatorname{Li}(\mathbb{R})$.) Here is a rather general result of this kind, due to Singer [1975]:

Theorem 1.2. *If $y \in \mathcal{C}^1$ satisfies a differential equation $y'P(y) = Q(y)$ where $P(Y)$ and $Q(Y)$ are polynomials over a Hardy field H and $P(y)$ is a unit in $\mathcal{C}$, then y generates a Hardy field $H\langle y\rangle = H(y, y')$ over H.*

Singer's theorem clearly does not extend to second order differential equations: the nonzero solutions of $y'' + y = 0$ in $\mathcal{C}^2$ do not belong to any Hardy field. The solutions in $\mathcal{C}^2$ of the differential equation

$$y'' + y = \mathrm{e}^{x^2} \tag{2}$$

form a two-dimensional affine space $y_0 + \mathbb{R}\sin x + \mathbb{R}\cos x$ over $\mathbb{R}$, with y_0 any particular solution. Boshernitzan [1987] proved that any of these continuum many solutions generates a Hardy field. Since no Hardy field can contain more than one solution, there are at least continuum many different maximal Hardy fields. By the above, each of them contains $\mathbb{R}$, is real closed, and closed under integration and exponentiation. *What more can we say about maximal Hardy fields?* To give an answer to this question, consider the following conjectures about Hardy fields H:

A. *For any differential polynomial $P(Y) \in H\{Y\} = H[Y, Y', Y'', \dots]$ and $f < g$ in H with $P(f) < 0 < P(g)$ there exists y in a Hardy field extension of H such that $f < y < g$ and $P(y) = 0$.*

B. *For any countable subsets $A < B$ in H there exists y in a Hardy field extension of H such that $A < y < B$.*

Conjecture A for $P \in H[Y, Y']$ holds by van den Dries [2000]. Conjecture A implies that *all maximal Hardy fields are elementarily equivalent* as we shall see in Section 2. Conjecture B was first raised as a question by Ehrlich [2012]. The conjectures together imply that, under CH, *all maximal Hardy fields are isomorphic* (the analogue of Corollary 1.1). We sketch a program to prove A and B in Section 3.

Transseries. Hardy [1910, p. 35] made the point that the LE-functions seem to cover all orders of infinity that occur naturally in mathematics. But he also suspected that the order of infinity of the compositional inverse of $(\log x)(\log\log x)$ differs from that of any LE-function (see Hardy [1912a]); this suspicion is correct. For a more revealing view of orders of infinity and a more comprehensive theory we need transseries. For example, transseries lead to an easy argument to confirm Hardy's suspicion (see van den Dries, Macintyre, and Marker [1997] and van der Hoeven [1997]). Here we focus on the field $\mathbb{T}$ of LE-series and in accordance with Aschenbrenner, van den Dries, and van der Hoeven [2017a], simply call its elements *transseries,* bearing in mind that many variants of formal series, such as those appearing in Schmeling [2001] (see Section 4 below), can also rightfully be called "transseries".

Transseries are formal series $f = \sum_{\mathfrak{m}} f_{\mathfrak{m}}\mathfrak{m}$ where the $f_{\mathfrak{m}}$ are real coefficients and the $\mathfrak{m}$ are "transmonomials" such as

$$x^r\ (r \in \mathbb{R}), \quad x^{-\log x}, \quad \mathrm{e}^{x^2\,\mathrm{e}^x}, \quad \mathrm{e}^{\mathrm{e}^x}.$$

One can get a sense by considering an example like

$$7\,\mathrm{e}^{\mathrm{e}^x+\mathrm{e}^{x/2}+\mathrm{e}^{x/4}+\cdots} - 3\,\mathrm{e}^{x^2} + 5x^{\sqrt{2}} - (\log x)^{\pi} + 42 + x^{-1} + x^{-2} + \cdots + \mathrm{e}^{-x}.$$

Here think of x as positive infinite: $x > \mathbb{R}$. The transmonomials in this series are arranged from left to right in decreasing order. The reversed order type of the set of transmonomials that occur in a given transseries can be any countable ordinal. (In the example above it is $\omega + 1$ because of the term e^{-x} at the end.) Formally, $\mathbb{T}$ is an ordered subfield of a Hahn field $\mathbb{R}[[G^{\mathrm{LE}}]]$ where G^{LE} is the ordered group of transmonomials (or LE-monomials). More generally, let $\mathfrak{M}$ be any (totally) ordered commutative group, multiplicatively written, the $\mathfrak{m} \in \mathfrak{M}$ being thought of as monomials, with the ordering denoted by $\preccurlyeq$. The Hahn field $\mathbb{R}[[\mathfrak{M}]]$ consists of the formal series $f = \sum_{\mathfrak{m}} f_{\mathfrak{m}}\mathfrak{m}$ with real coefficients $f_{\mathfrak{m}}$ whose support $\operatorname{supp} f := \{\mathfrak{m} \in \mathfrak{M} : f_{\mathfrak{m}} \neq 0\}$ is *well-based*, that is, well-ordered in the reversed ordering $\succcurlyeq$ of $\mathfrak{M}$. Addition and multiplication of these Hahn series works just as for ordinary power series, and the ordering of $\mathbb{R}[[\mathfrak{M}]]$ is determined by declaring a nonzero Hahn series to be positive if its leading coefficient is positive (so the series above, with leading coefficient 7, is positive). Both $\mathbb{R}[[G^{\mathrm{LE}}]]$ and its ordered subfield $\mathbb{T}$ are real closed. Informally, each transseries is obtained, starting with the powers x^r $(r \in \mathbb{R})$, by applying the following operations finitely many times:

1. multiplication with real numbers;
2. infinite summation in $\mathbb{R}[[G^{\mathrm{LE}}]]$;
3. exponentiation and taking logarithms of positive transseries.

To elaborate on (2), a family $(f_i)_{i\in I}$ in $\mathbb{R}[[\mathfrak{M}]]$ is said to be *summable* if for each $\mathfrak{m}$ there are only finitely many $i \in I$ with $\mathfrak{m} \in \operatorname{supp} f_i$, and $\bigcup_{i\in I} \operatorname{supp} f_i$ is well-based; in this case we define the *sum* $f = \sum_{i\in I} f_i \in \mathbb{R}[[\mathfrak{M}]]$ of this family by $f_{\mathfrak{m}} = \sum_{i\in I}(f_i)_{\mathfrak{m}}$ for each $\mathfrak{m}$. One can develop a "strong" linear algebra for this notion of "strong" (infinite) summation (see van der Hoeven [2006] and Schmeling [2001]). As to (3), it may be instructive to see how to exponentiate a transseries f: decompose f as $f = g + c + \varepsilon$ where $g := \sum_{\mathfrak{m}\succ 1} f_{\mathfrak{m}}\mathfrak{m}$ is the infinite part of f, $c := f_1$ is its constant term, and ε its infinitesimal part (in our example $c = 42$ and $\varepsilon = x^{-1} + x^{-2} + \cdots + e^{-x}$); then

$$e^f = e^g \cdot e^c \cdot \sum_n \frac{\varepsilon^n}{n!}$$

where $e^g \in \mathfrak{M}$ is a transmonomial, and $e^c \in \mathbb{R}$, $\sum_n \frac{\varepsilon^n}{n!} \in \mathbb{R}[[G^{\mathrm{LE}}]]$ have their usual meaning. The story with logarithms is a bit different: taking logarithms may also create transmonomials, such as $\log x$, $\log\log x$, etc.

The formal definition of $\mathbb{T}$ is inductive and somewhat lengthy; see van den Dries, Macintyre, and Marker [2001], Edgar [2010], and van der Hoeven [2006] for detailed expositions, or Aschenbrenner, van den Dries, and van der Hoeven [2017a, Appendix A] for a summary. We only note here that by virtue of the construction of $\mathbb{T}$, series like $\frac{1}{x} + \frac{1}{e^x} + \frac{1}{e^{e^x}} + \cdots$ or $\frac{1}{x} + \frac{1}{x\log x} + \frac{1}{x\log x\log\log x} + \cdots$ (involving "nested" exponentials or logarithms of unbounded depth), though they are legitimate elements of $\mathbb{R}[[G^{\mathrm{LE}}]]$, do not appear in $\mathbb{T}$; moreover, the sequence $x, e^x, e^{e^x}, \dots$ is cofinal in $\mathbb{T}$, and the sequence $x, \log x, \log\log x, \dots$ is coinitial in the set $\{f \in \mathbb{T} : f > \mathbb{R}\}$. The map $f \mapsto e^f$ is an isomorphism of the ordered additive group of $\mathbb{T}$ onto its multiplicative group of positive elements, with inverse $g \mapsto \log g$. As an ordered exponential field, $\mathbb{T}$ turns out to be an elementary extension of $\mathbb{R}_{\exp}$ (see van den Dries, Macintyre, and Marker [1997]).

Transseries can be differentiated termwise; for instance, $\left(\sum_n n!\frac{e^x}{x^{n+1}}\right)' = \frac{e^x}{x}$. We obtain a derivation $f \mapsto f'$ on the field $\mathbb{T}$ with constant field $\{f \in \mathbb{T} : f' = 0\} = \mathbb{R}$ and satisfying $(\exp f)' = f'\exp f$ and $(\log g)' = g'/g$ for $f, g \in \mathbb{T}$, $g > 0$. Moreover, each $f \in \mathbb{T}$ has an antiderivative in $\mathbb{T}$, that is, $f = g'$ for some $g \in \mathbb{T}$. As in Hardy fields, $f > \mathbb{R} \Rightarrow f' > 0$, for transseries f. We also have a dominance relation on $\mathbb{T}$: for $f, g \in \mathbb{T}$ we set

$$\begin{aligned} f \preccurlyeq g \ &:\Longleftrightarrow\ |f| \leqslant c|g| \text{ for some } c \in \mathbb{R}^{>0} \\ &\Longleftrightarrow\ (\text{leading transmonomial of } f) \preccurlyeq (\text{leading transmonomial of } g), \end{aligned}$$

and as in Hardy fields we declare $f \asymp g :\Longleftrightarrow f \preccurlyeq g$ and $g \preccurlyeq f$, as well as $f \prec g :\Longleftrightarrow f \preccurlyeq g$ and $g \not\preccurlyeq f$. As in Hardy fields we can also differentiate and integrate asymptotic relations: for $0 \neq f, g \not\asymp 1$ in $\mathbb{T}$ we have $f \preccurlyeq g$ iff $f' \preccurlyeq g'$.

Hardy's ordered exponential field of (germs of) logarithmic-exponential functions embeds uniquely into $\mathbb{T}$ so as to preserve real constants and to send the germ x to the transseries x; this embedding also preserves the derivation. However, the field of LE-series enjoys many closure properties that the field of LE-functions lacks. For instance, $\mathbb{T}$ is not only closed under exponentiation and integration, but also comes with a natural operation of composition: for $f, g \in \mathbb{T}$ with $g > \mathbb{R}$ we can substitute g for x in $f = f(x)$ to obtain $f \circ g = f(g(x))$. The Chain Rule holds: $(f \circ g)' = (f' \circ g) \cdot g'$. Every $g > \mathbb{R}$ has a compositional inverse in $\mathbb{T}$: a transseries $f > \mathbb{R}$ with $f \circ g = g \circ f = x$. As shown in van der Hoeven [2006], a Newton diagram method can be used to solve any "feasible" algebraic differential equation in $\mathbb{T}$ (where the meaning of feasible can be made explicit).

Thus it is not surprising that soon after the introduction of $\mathbb{T}$ the idea emerged that it should play the role of a *universal domain* (akin to Weil's use of this term in algebraic geometry) for asymptotic differential algebra: that it *is truly the algebra-from-which-one-can-never-exit and that it marks an almost impassable horizon for "ordered analysis"*, as Écalle [1992, p. 148] put it. Model theory provides a language to make such an intuition precise, as we explain in our survey Aschenbrenner, van den Dries, and van der Hoeven [2013] where we sketched a program to establish the basic model-theoretic properties of $\mathbb{T}$, carried out in Aschenbrenner, van den Dries, and van der Hoeven [2017a]. Next we briefly discuss our main results from Aschenbrenner, van den Dries, and van der Hoeven [ibid.].

2 H-Fields

We shall consider $\mathbb{T}$ as an $\mathcal{L}$-structure where the language $\mathcal{L}$ has the primitives 0, 1, $+$, $-$, $\cdot$, ∂ (derivation), $\leqslant$ (ordering), $\preccurlyeq$ (dominance). More generally, let K be any ordered differential field with constant field $C = \{f \in K : f' = 0\}$. This yields a dominance relation $\preccurlyeq$ on K by

$$f \preccurlyeq g \quad :\Longleftrightarrow \quad |f| \leqslant c|g| \text{ for some positive } c \in C$$

and we view K accordingly as an $\mathcal{L}$-structure. The convex hull of C in K is the valuation ring $\mathcal{O} = \{f \in K : f \preccurlyeq 1\}$ of K, with its maximal ideal $\scriptstyle\mathcal{O} := \{f \in K : f \prec 1\}$ of infinitesimals.

Definition 2.1. *An* H-field *is an ordered differential field* K *such that* (*with the notations above*), $\mathcal{O} = C + \scriptstyle\mathcal{O}$, *and for all* $f \in K$ *we have:* $f > C \Rightarrow f' > 0$.

Examples include all Hardy fields that contain $\mathbb{R}$, and all ordered differential subfields of $\mathbb{T}$ that contain $\mathbb{R}$. In particular, $\mathbb{T}$ is an H-field, but $\mathbb{T}$ has further basic elementary properties that do not follow from this: its derivation is small, and it is Liouville closed.

An H-field K is said to have *small derivation* if it satisfies $f \prec 1 \Rightarrow f' \prec 1$, and to be *Liouville closed* if it is real closed and for every $f \in K$ there are $g, h \in K$, $h \neq 0$, such that $g' = f$ and $h' = hf$. Each Hardy field H has small derivation, and $\operatorname{Li}(H)$ is Liouville closed.

Inspired by the familiar characterization of real closed ordered fields via the intermediate value property for one-variable polynomial functions, we say that an H-field K has the *Intermediate Value Property* (IVP) if for all differential polynomials $P(Y) \in K\{Y\}$ and all $f < g$ in K with $P(f) < 0 < P(g)$ there is some $y \in K$ with $f < y < g$ and $P(y) = 0$. van der Hoeven showed that a certain variant of $\mathbb{T}$, namely its H-subfield of gridbased transseries, has IVP; see van der Hoeven [2002].

Theorem 2.2. *The $\mathfrak{L}$-theory of $\mathbb{T}$ is completely axiomatized by the requirements: being an H-field with small derivation; being Liouville closed; and having IVP.*

Actually, IVP is a bit of an afterthought: in Aschenbrenner, van den Dries, and van der Hoeven [2017a] we use other (but equivalent) axioms that will be detailed below. We mention the above variant for expository reasons and since it explains why Conjecture A from Section 1 yields that all maximal Hardy fields are elementarily equivalent. Let us define an *H-closed field* to be an H-field that is Liouville closed and has the IVP. All H-fields embed into H-closed fields, and the latter are exactly the existentially closed H-fields. Thus:

Theorem 2.3. *The theory of H-closed fields is model complete.*

Here is an unexpected byproduct of our proof of this theorem:

Corollary 2.4. *H-closed fields have no proper differentially algebraic H-field extensions with the same constant field.*

IVP refers to the ordering, but the valuation given by $\preccurlyeq$ is more robust and more useful. IVP comes from two more fundamental properties: ω-*freeness* and *newtonianity* (a differential version of henselianity). These concepts make sense for any differential field with a suitable dominance relation $\preccurlyeq$ in which the equivalence $f \preccurlyeq g \iff f' \preccurlyeq g'$ holds for $0 \neq f, g \prec 1$.

To give an inkling of these somewhat technical notions, let K be an H-field and assume that for every $\phi \in K^\times$ for which the derivation $\phi\partial$ is small (that is, $\phi\partial\mathcal{O} \subseteq \mathcal{O}$), there exists $\phi_1 \prec \phi$ in $K^\times$ such that $\phi_1\partial$ is small. (This assumption is satisfied for Liouville closed H-fields.) Let $P(Y) \in K\{Y\}^{\neq}$. We wish to understand how the function $y \mapsto P(y)$ behaves for $y \preccurlyeq 1$. It turns out that this function only reveals its true colors after rewriting P in terms of a derivation $\phi\partial$ with suitable $\phi \in K^\times$.

Indeed, this rewritten P has the form $a \cdot (N + R)$ with $a \in K^\times$ and where $N(Y) \in C\{Y\}^{\neq}$ is independent of ϕ for sufficiently small $\phi \in K^\times$ with respect to $\preccurlyeq$, subject to $\phi\partial$

being small, and where the coefficients of $R(Y)$ are infinitesimal. We call N the *Newton polynomial* of P. Now K is said to be ω*-free* if for all P as above its Newton polynomial has the form $A(Y)\cdot(Y')^n$ for some $A \in C[Y]$ and some n. We say that K is *newtonian* if for all P as above with $N(P)$ of degree 1 we have $P(y) = 0$ for some $y \in \mathcal{O}$. For H-fields, IVP $\Longrightarrow$ ω-free and newtonian; for Liouville closed H-fields, the converse also holds.

Our main result in Aschenbrenner, van den Dries, and van der Hoeven [ibid.] refines Theorem 2.3 by giving quantifier elimination for the theory of H-closed fields in the language $\mathfrak{L}$ above augmented by an additional unary function symbol ι and two extra unary predicates Λ and Ω. These have defining axioms in terms of the other primitives. Their interpretations in $\mathbb{T}$ are as follows: $\iota(f) = 1/f$ if $f \neq 0$, $\iota(0) = 0$, and with $\ell_0 := x$, $\ell_{n+1} := \log \ell_n$,

$$\Lambda(f) \iff f < \lambda_n := \tfrac{1}{\ell_0} + \tfrac{1}{\ell_0\ell_1} + \cdots + \tfrac{1}{\ell_0\cdots\ell_n} \text{ for some } n,$$

$$\Omega(f) \iff f < \omega_n := \tfrac{1}{(\ell_0)^2} + \tfrac{1}{(\ell_0\ell_1)^2} + \cdots + \tfrac{1}{(\ell_0\cdots\ell_n)^2} \text{ for some } n.$$

Thus Λ and Ω define downward closed subsets of $\mathbb{T}$. The sequence (ω_n) also appears in classical non-oscillation theorems for second-order linear differential equations. The ω-freeness of $\mathbb{T}$ reflects the fact that (ω_n) has no pseudolimit in the valued field $\mathbb{T}$. Here are some applications of this quantifier elimination:

Corollary 2.5.

(1) *"O-minimality at infinity": if $S \subseteq \mathbb{T}$ is definable, then for some $f \in \mathbb{T}$ we either have $g \in S$ for all $g > f$ in $\mathbb{T}$ or $g \notin S$ for all $g > f$ in $\mathbb{T}$.*

(2) *All subsets of $\mathbb{R}^n$ definable in $\mathbb{T}$ are semialgebraic.*

Corollaries 2.4 and 2.5 are the departure point for developing a notion of (differential-algebraic) dimension for definable sets in $\mathbb{T}$; see Aschenbrenner, van den Dries, and van der Hoeven [2017b].

The results reported on above make us confident that the category of H-fields is the right setting for asymptotic differential algebra. To solidify this impression we return to the motivating examples—Hardy fields, ordered differential fields of transseries, and surreal numbers—and consider how they are related. We start with Hardy fields, which historically came first.

3 H-Field Elements as Germs

After Theorem 1.2 and Boshernitzan [1982, 1987], the first substantial "Hardy field" result on more general differential equations was obtained by van der Hoeven [2009]. In what follows we use "d-algebraic" to mean "differentially algebraic" and "d-transcendental" to mean "differentially transcendental".

Theorem 3.1. *The differential subfield* $\mathbb{T}^{\mathrm{da}}$ *of* $\mathbb{T}$ *whose elements are the* d-*algebraic transseries is isomorphic over* $\mathbb{R}$ *to a Hardy field.*

The proof of this theorem is in the spirit of model theory, iteratively extending by a single d-algebraic transseries. The most difficult case (immediate extensions) is handled through careful construction of suitable solutions as convergent series of iterated integrals. We are currently trying to generalize Theorem 3.1 to d-algebraic extensions of arbitrary Hardy fields. Here is our plan:

Theorem 3.2. *Every Hardy field has an* ω*-free Hardy field extension.*

Theorem 3.3 (in progress)**.** *Every* ω*-free Hardy field has a newtonian* d-*algebraic Hardy field extension.*

These two theorems, when established, imply that all maximal Hardy fields are H-closed. Hence (by Theorem 2.2) they will all be elementarily equivalent to $\mathbb{T}$, and since H-closed fields have the IVP, Conjecture A from Section 1 will follow.

In order to get an even better grasp on the structure of maximal Hardy fields, we also need to understand how to adjoin d-transcendental germs to Hardy fields. An example of this situation is given by d-transcendental series such as $\sum_n n!!x^{-n}$. By an old result by É. Borel [1895] every formal power series $\sum_n a_n t^n$ over $\mathbb{R}$ is the Taylor series at 0 of a $\mathcal{C}^\infty$-function f on $\mathbb{R}$; then $\sum_n a_n x^{-n}$ is an asymptotic expansion of the function $f(x^{-1})$ at $+\infty$, and it is easy to show that if this series is d-transcendental, then the germ at $+\infty$ of this function does generate a Hardy field. Here is a far-reaching generalization:

Theorem 3.4 (in progress)**.** *Every pseudocauchy sequence* (y_n) *in a Hardy field* H *has a pseudolimit in some Hardy field extension of* H.

The proof of this for H-closed $H \supseteq \mathbb{R}$ relies heavily on results from Aschenbrenner, van den Dries, and van der Hoeven [2017a], using also intricate glueing techniques. For extensions that increase the value group, we need very different constructions. If succesful, these constructions in combination with Theorem 3.4 will lead to a proof of Conjecture B from Section 1:

Theorem 3.5 (in progress)**.** *For any countable subsets* $A < B$ *of a Hardy field* H *there exists an element* y *in a Hardy field extension of* H *with* $A < y < B$.

The case $H \subseteq \mathcal{C}^\infty$, $B = \emptyset$ was already dealt with by Sjödin [1971]. The various "theorems in progress" together with results from Aschenbrenner, van den Dries, and van der Hoeven [2017a] imply that any maximal Hardy fields H_1 and H_2 are back-and-forth equivalent, which is considerably stronger than H_1 and H_2 being elementarily equivalent. It implies for example

> *Under* CH *all maximal Hardy fields are isomorphic.*

This would be the Hardy field analogue of Corollary 1.1. (In contrast to maximal Hausdorff fields, however, maximal Hardy fields cannot be $\aleph_1$-saturated, since their constant field is $\mathbb{R}$.) When we submitted this manuscript, we had finished the proof of Theorem 3.2, and also the proof of Theorem 3.4 in the relevant H-closed case.

Related problems. Some authors (such as Sjödin [1971]) prefer to consider only Hardy fields contained in $\mathcal{C}^\infty$. Theorem 3.2 and our partial result for Theorem 3.4 go through in the $\mathcal{C}^\infty$-setting. All the above "theorems in progress" are plausible in that setting.

What about real analytic Hardy fields (Hardy fields contained in the subring $\mathcal{C}^\omega$ of $\mathcal{C}$ consisting of all real analytic germs)? In that setting Theorem 3.2 goes through. Any d-algebraic Hardy field extension of a real analytic Hardy field is itself real analytic, and so Theorem 3.3 (in progress) will hold in that setting as well. However, our glueing technique employed in the proof of Theorem 3.4 doesn't work there.

Kneser [1949] obtained a real analytic solution E at infinity to the functional equation $E(x+1) = \exp E(x)$. It grows faster than any finite iteration of the exponential function, and generates a Hardy field. See Boshernitzan [1986] for results of this kind, and a proof that Theorem 3.5 holds for $B = \varnothing$ in the real analytic setting. So in this context we also have an abundant supply of Hardy fields.

Similar issues arise for germs of quasi-analytic and "cohesive" functions of Écalle [1992]. These classes of functions are somewhat more flexible than the class of real analytic functions. For instance, the series $x^{-1} + \mathrm{e}^{-x} + \mathrm{e}^{-\mathrm{e}^x} + \cdots$ converges uniformly for $x > 1$ to a cohesive function that is not real analytic.

Accelero-summation. The definition of a Hardy field ensures that the differential field operations never introduce oscillatory behavior. Does this behavior persist for operations such as composition or various integral transforms? In this connection we note that the Hardy field $H(\tilde{\mathbb{R}})$ associated to an o-minimal expansion $\tilde{\mathbb{R}}$ of the field of reals is always closed under composition (see Miller [2012]).

To illustrate the problem with composition, let α be a real number > 1 and let $y_0 \in \mathcal{C}^2$ be a solution to (2). Then $z_0 := y_0(\alpha x)$ satisfies the equation

$$\alpha^{-2} z'' + z \;=\; \mathrm{e}^{\alpha^2 x^2}. \tag{3}$$

It can be shown that $\{y_0 + \sin x, z_0\}$ generates a Hardy field, but it is clear that no Hardy field containing both $y_0 + \sin x$ and z_0 can be closed under composition.

Adjoining solutions to (2) and (3) "one by one" as in the proof of Theorem 3.1 will not prevent the resulting Hardy fields to contain both $y_0 + \sin x$ and z_0. In order to obtain closure under composition we therefore need an alternative device. Écalle's theory of *accelero-summation* (Écalle [1992]) is much more than that. Vastly extending Borel's summation method for divergent series (Borel [1899]), it associates to each *accelero-summable* transseries an *analyzable* function. In this way many non-oscillating real-valued functions that arise naturally (e.g., as solutions of algebraic differential equations) can be represented faithfully by transseries. This leads us to conjecture an improvement on Theorem 3.1:

Conjecture 3.6. *Consider the real accelero-summation process where we systematically use the organic average whenever we encounter singularities on the positive real axis. This yields a composition-preserving H-field isomorphism from $\mathbb{T}^{\mathrm{da}}$ onto a Hardy field contained in $\mathcal{C}^{\omega}$.*

There is little doubt that this holds. The main difficulty here is that a full proof will involve many tools forged by Écalle in connection with accelero-summation, such as resurgent functions, well-behaved averages, cohesive functions, etc., with some of these tools requiring further elaboration; see also Costin [2009] and Menous [1999].

The current theory of accelero-summation only sums transseries with coefficients in $\mathbb{R}$. Thus it is not clear how to generalize Conjecture 3.6 in the direction of Theorem 3.3. Such a generalization might require introducing transseries over a Hardy field H with suitable additional structure, as well as a corresponding theory of accelero-summation over H for such transseries. In particular, elements of H should be accelero-summable over H in this theory, by construction.

4 H-Field Elements as Generalized Transseries

Next we discuss when H-fields embed into differential fields of formal series. A classical embedding theorem of this type is due to Krull [1932]: any valued field has a spherically complete immediate extension. As a consequence, any real closed field containing $\mathbb{R}$ is isomorphic over $\mathbb{R}$ to a subfield of a Hahn field $\mathbb{R}[[\mathfrak{M}]]$ with divisible monomial group $\mathfrak{M}$, such that the subfield contains $\mathbb{R}(\mathfrak{M})$. We recently proved an analogue of this theorem for valued differential fields; see Aschenbrenner, van den Dries, and van der Hoeven [2017c]. Here a *valued differential field* is a valued field of equicharacteristic zero equipped with a derivation that is continuous with respect to the valuation topology.

Theorem 4.1. *Every valued differential field has a spherically complete immediate extension.*

For a real closed H-field K with constant field C this theorem gives a Hahn field $\widehat{K} = C[[\mathfrak{M}]]$ with a derivation ∂ on $\widehat{K}$ making it an H-field with constant field C such that K is isomorphic over C to an H-subfield of $\hat{K}$ that contains $C(\mathfrak{M})$. A shortcoming of this result is that there is no guarantee that ∂ preserves infinite summation. In contrast, the derivation of $\mathbb{T}$ is *strong* (does preserve infinite summation). An abstract framework for even more general notions of transseries is due to van der Hoeven and his former student Schmeling [2001].

Fields of transseries. To explain this, consider an (ordered) Hahn field $\mathbb{R}[[\mathfrak{M}]]$ with a partially defined function exp obeying the usual rules of exponentiation; see van der Hoeven [2006, Section 4.1] for details. In particular, exp has a partially defined inverse function log. We say that $\mathbb{R}[[\mathfrak{M}]]$ is a *field of transseries* if the following conditions hold:

(T1) the domain of the function log is $\mathbb{R}[[\mathfrak{M}]]^{>0}$;

(T2) for each $\mathfrak{m} \in \mathfrak{M}$ and $\mathfrak{n} \in \operatorname{supp} \log \mathfrak{m}$ we have $\mathfrak{n} \succ 1$;

(T3) $\log(1+\varepsilon) = \varepsilon - \frac{1}{2}\varepsilon^2 + \frac{1}{3}\varepsilon^3 + \cdots$ for all $\varepsilon \prec 1$ in $\mathbb{R}[[\mathfrak{M}]]$; and

(T4) for every sequence $(\mathfrak{m}_n)$ in $\mathfrak{M}$ with $\mathfrak{m}_{n+1} \in \operatorname{supp} \log \mathfrak{m}_n$ for all n, there exists an index n_0 such that for all $n \geqslant n_0$ and all $\mathfrak{n} \in \operatorname{supp} \log \mathfrak{m}_n$, we have $\mathfrak{n} \succcurlyeq \mathfrak{m}_{n+1}$ and $(\log \mathfrak{m}_n)_{\mathfrak{m}_{n+1}} = \pm 1$.

The first three axioms record basic facts from the standard construction of transseries. The fourth axiom is more intricate and puts limits on the kind of “nested transseries” that are allowed. Nested transseries such as

$$y = \sqrt{x} + \mathrm{e}^{\sqrt{\log x} + \mathrm{e}^{\sqrt{\log\log x} + \mathrm{e}^{\cdots}}} \tag{4}$$

are naturally encountered as solutions of functional equations, in this case

$$y(x) = \sqrt{x} + \mathrm{e}^{y(\log x)}. \tag{5}$$

Axiom (T4) does allow nested transseries as in (4), but excludes series like

$$u = \sqrt{x} + \mathrm{e}^{\sqrt{\log x} + \mathrm{e}^{\sqrt{\log\log x} + \mathrm{e}^{\cdots} + \log\log\log x} + \log\log x} + \log x,$$

which solves the functional equation $u(x) = \sqrt{x} + \mathrm{e}^{u(\log x)} + \log x$; in some sense, u is a perturbation of the solution y in (4) to the equation (5).

In his thesis Schmeling [2001] shows how to extend a given field of transseries $K = \mathbb{R}[[\mathfrak{M}]]$ with new exponentials and nested transseries like (4), and if K also comes with a strong derivation, how to extend this derivation as well. Again, (T4) is crucial for this task: naive termwise differentiation leads to a huge infinite sum that turns out to be summable by (T4). A *transserial derivation* is a strong derivation on K such that nested transseries are differentiated in this way. Such a transserial derivation is uniquely determined by its values on the *log-atomic* elements: those $\lambda \in K$ for which $\lambda, \log\lambda, \log\log\lambda, \ldots$ are all transmonomials in $\mathfrak{M}$.

We can now state a transserial analogue of Krull's theorem. This analogue is a consequence of Theorem 5.3 below, proved in Aschenbrenner, van den Dries, and van der Hoeven [2015].

Theorem 4.2. *Every H-field with small derivation and constant field $\mathbb{R}$ can be embedded over $\mathbb{R}$ into a field of transseries with transserial derivation.*

For simplicity, we restricted ourselves to transseries over $\mathbb{R}$. The theory naturally generalizes to transseries over ordered exponential fields (see van der Hoeven [2006] and Schmeling [2001]) and it should be possible to extend Theorem 4.2 likewise.

Hyperseries. Besides derivations, one can also define a notion of composition for generalized transseries (see van der Hoeven [1997] and Schmeling [2001]). Whereas certain functional equations such as (5) can still be solved using nested transseries, solving the equation $E(x+1) = \exp E(x)$ where $E(x)$ is the unknown, requires extending $\mathbb{T}$ to a field of transseries with composition containing an element $E(x) = \exp_\omega x > \mathbb{T}$, called the *iterator* of $\exp x$. Its compositional inverse $\log_\omega x$ should then satisfy $\log_\omega \log x = (\log_\omega x) - 1$, providing us with a primitive for $(x \log x \log_2 x \cdots)^{-1}$:

$$\log_\omega x = \int \frac{\mathrm{d}x}{x \log x \log_2 x \cdots}.$$

It is convenient to start with iterated logarithms rather than iterated exponentials, and to introduce transfinite iterators $\log_\alpha x$ recursively using

$$\log_\alpha x = \int \frac{\mathrm{d}x}{\prod_{\beta<\alpha} \log_\beta x} \qquad (\alpha \text{ any ordinal}).$$

By Écalle [1992] the iterators $\log_\alpha x$ with $\alpha < \omega^\omega$ and their compositional inverses $\exp_\alpha x$ suffice to resolve all pure composition equations of the form

$$f^{\circ k_1} \circ \phi_1 \circ \cdots \circ f^{\circ k_n} \circ \phi_n = x \quad \text{where } \phi_1, \ldots, \phi_n \in \mathbb{T} \text{ and } k_1, \ldots, k_n \in \mathbb{N}.$$

The resolution of more complicated functional equations involving differentiation and composition requires the introduction of fields of *hyperseries*: besides exponentials and logarithms, hyperseries are allowed to contain iterators $\exp_\alpha x$ and $\log_\alpha x$ of any strength α.

For $\alpha < \omega^\omega$, the necessary constructions were carried out in Schmeling [2001]. The ultimate objective is to construct a field **Hy** of hyperseries as a proper class, similar to the field of surreal numbers, endow it with its canonical derivation and composition, and establish the following:

Conjecture 4.3. *Let Φ be any partial function from* **Hy** *into itself, constructed from elements in* **Hy**, *using the field operations, differentiation and composition. Let $f < g$ be hyperseries in* **Hy** *such that Φ is defined on the closed interval $[f, g]$ and $\Phi(f)\Phi(g) < 0$. Then for some $y \in$* **Hy** *we have $\Phi(y) = 0$ and $f < y < g$.*

One might then also consider H-fields with an additional composition operator and try to prove that these structures can always be embedded into **Hy**.

5 Growth Rates as Numbers

Turning to surreal numbers, how do they fit into asymptotic differential algebra?

The H-field of surreal numbers. The totality **No** of surreal numbers is not a set but a proper class: a surreal $a \in$ **No** is uniquely represented by a transfinite sign sequence $(a_\lambda)_{\lambda<\ell(a)} \in \{-,+\}^{\ell(a)}$ where $\ell(a)$ is an ordinal, called the *length* of a; a surreal b is said to be *simpler than* a (notation: $b <_s a$) if the sign sequence of b is a proper initial segment of that of a. Besides the (partial) ordering $<_s$, **No** also carries a natural (total) lexicographic ordering $<$. For any *sets* $L < R$ of surreals there is a unique simplest surreal a with $L < a < R$; this a is denoted by $\{L \mid R\}$ and called the *simplest* or *earliest* surreal between L and R. In particular, $a = \{L_a \mid R_a\}$ for any $a \in$ **No**, where $L_a := \{b <_s a : b < a\}$ and $R_a = \{b <_s a : b > a\}$. We let a^L range over elements of L_a, and a^R over elements of R_a.

A rather magical property of surreal numbers is that various operations have natural inductive definitions. For instance, we have ring operations given by

$$\begin{aligned} a+b &:= \{a^L+b, a+b^L \mid a^R+b, a+b^R\} \\ ab &:= \{a^Lb+ab^L-a^Lb^L, a^Rb+ab^R-a^Rb^R \mid \\ &\qquad a^Lb+ab^R-a^Lb^R, a^Rb+ab^L-a^Rb^L\}. \end{aligned}$$

Remarkably, these operations make **No** into a real closed field with $<$ as its field ordering and with $\mathbb{R}$ uniquely embedded as an initial subfield. (A set $A \subseteq$ **No** is said to be *initial* if for all $a \in A$ all $b <_s a$ are also in A.)

Can we use such magical recursions to introduce other reasonable operations? Exponentiation was dealt with by Gonshor [1986]. But it remained long open how to define a "good" derivation ∂ on **No** such that $\partial(\omega) = 1$. (An ordinal α is identified with the surreal of length α whose sign sequence has just plus signs.) A positive answer was given recently by Berarducci and Mantova [2018]. Their construction goes in two parts. They first analyze **No** as an exponential field, and show that it is basically a field of transseries in the sense of Section 4. A transserial derivation on **No** is determined by its values at log-atomic elements. There is some flexibility here, but Berarducci and Mantova [ibid.] present a "simplest" way to choose these derivatives. Most important, that choice indeed leads to a derivation ∂_{BM} on **No**. In addition:

Theorem 5.1 (Berarducci and Mantova [ibid.]). *The derivation ∂_{BM} is transserial and makes* **No** *a Liouville closed H-field with constant field $\mathbb{R}$.*

This result was further strengthened in Aschenbrenner, van den Dries, and van der Hoeven [2015], using key results from Aschenbrenner, van den Dries, and van der Hoeven [2017a]:

Theorem 5.2. **No** *with the derivation ∂_{BM} is an H-closed field.*

Embedding H-fields into No. In the remainder of this section we consider **No** as equipped with the derivation ∂_{BM}, although Theorems 5.1 and 5.2 and much of what follows hold for other transserial derivations. Returning to our main topic of embedding H-fields into specific H-fields such as **No**, we also proved the following in Aschenbrenner, van den Dries, and van der Hoeven [2015]:

Theorem 5.3. *Every H-field with small derivation and constant field $\mathbb{R}$ can be embedded as an ordered differential field into* **No**.

How "nice" can we take the embeddings in Theorem 5.3? For instance, when can we arrange the image of the embedding to be initial? The image of the natural embedding $\mathbb{T} \to \mathbf{No}$ is indeed initial, as has been shown by Elliot Kaplan.

For further discussion it is convenient to introduce, given an ordinal α, the *set*

$$\mathbf{No}(\alpha) := \{a \in \mathbf{No} : \ell(a) < \alpha\}.$$

It turns out that for uncountable cardinals κ, $\mathbf{No}(\kappa)$ is closed under the differential field operations, and in Aschenbrenner, van den Dries, and van der Hoeven [ibid.] we also show:

Theorem 5.4. *The H-subfield $\mathbf{No}(\kappa)$ of* **No** *is an elementary submodel of* **No**.

In particular, the H-field $\mathbf{No}(\omega_1)$ of surreal numbers of countable length is an elementary submodel of **No**. It has the η_1-property: for any countable subsets $A < B$ of $\mathbf{No}(\omega_1)$ there exists $y \in \mathbf{No}(\omega_1)$ with $A < y < B$. This fact and the various "theorems in progress" from Section 3 imply:

Under CH *all maximal Hardy fields are isomorphic to* $\mathbf{No}(\omega_1)$.

This would be an analogue of Ehrlich's observation about maximal Hausdorff fields.

Hyperseries as numbers and *vice versa*. The similarities in the constructions of the field of hyperseries **Hy** and the field of surreal numbers **No** led van der Hoeven [2006, p. 6] to the following:

Conjecture 5.5. *There is a natural isomorphism between* **Hy** *and* **No** *that associates to any hyperseries* $f(x) \in \mathbf{Hy}$ *its value* $f(\omega) \in \mathbf{No}$.

The problem is to make sense of the value of a hyperseries at ω. Thanks to Gonshor's exponential function, it is clear how to evaluate ordinary transseries at ω. The difficulties start as soon as we wish to represent surreal numbers that are not of the form $f(\omega)$ with $f(x)$ an ordinary transseries. That is where the iterators $\exp_\omega$ and $\log_\omega$ come into play:

$$\begin{aligned}
\exp_\omega \omega &:= \{\omega, \exp \omega, \exp_2 \omega, \ldots \mid \} \\
\log_\omega \omega &:= \{ \mathbb{R} \mid \ldots, \log_2 \omega, \log \omega, \omega\} \\
\exp_{1/2} \omega &:= \exp_\omega \left(\log_\omega \left(\omega + \tfrac{1}{2}\right)\right) \\
&:= \left\{ \omega^2, \exp \log^2 \omega, \exp_2 \log_2^2 \omega, \ldots \,\middle|\, \ldots, \exp_2 \sqrt{\log \omega}, \exp \sqrt{\omega} \right\}
\end{aligned}$$

The intuition behind Conjecture 5.5 is that all "holes in **No** can be filled" using suitable nested hyperseries and suitable iterators of exp and log. It reconciles two *a priori* very different types of infinities: on the one hand, we have growth orders corresponding to smooth functional behavior; on the other side, we have numbers. Being able to switch between functions (more precisely: formal series acting as functions) and numbers, we may also transport any available structure in both directions: we immediately obtain a canonical derivation ∂_c (with constant field $\mathbb{R}$) and composition $\circ_c$ on **No**, as well as a notion of simplicity on **Hy**.

Does the derivation ∂_{BM} coincide with the canonical derivation ∂_c induced by the conjectured isomorphism? A key observation is that any derivation ∂ on **No** with a distinguished right inverse ∂^{-1} naturally gives rise to a definition of $\log_\omega$:

$$\log_\omega a := \partial^{-1}(\partial a \log'_\omega a) \quad \text{where}$$
$$\log'_\omega a := 1 \Big/ \prod_n \log_n a \qquad (a \in \mathbf{No},\ a > \mathbb{R}).$$

(For a family (a_i) of positive surreals, $\prod_i a_i := \exp \sum_i \log a_i$ if $\sum_i \log a_i$ is defined.) Since ∂_{BM} is transserial, it does admit a distinguished right inverse ∂_{BM}^{-1}. According to

Berarducci and Mantova [2018, Remark 6.8], $\partial_{\mathrm{BM}}\lambda = 1/\log'_\omega \lambda$ for log-atomic λ with $\lambda > \exp_n \omega$ for all n. For $\lambda = \exp_\omega \omega$ and setting $\exp'_\omega(a) := \prod_n \log_n \exp_\omega a$ for $a \in \mathbf{No}^{>0}$, this yields $\partial_{\mathrm{BM}}\lambda = \exp'_\omega \omega$, which is also the value we expect for $\partial_{\mathrm{c}}\lambda$. However, for $\lambda = \exp_\omega(\exp_\omega \omega)$ we get $\partial_{\mathrm{BM}}\lambda = \exp'_\omega(\exp_\omega \omega)$ whereas we expect $\partial_{\mathrm{c}}\lambda = (\exp'_\omega \omega)\cdot \exp'_\omega(\exp_\omega \omega)$. Thus the "simplest" derivation ∂_{BM} making $\mathbf{No}$ an H-field probably does *not* coincide with the ultimately "correct" derivation ∂_{c} on $\mathbf{No}$. Berarducci and Mantova [2017] use similar considerations to conclude that ∂_{BM} is incompatible with any reasonable notion of composition for surreal numbers.

The surreal numbers from a model theoretic perspective. We conclude with speculations motivated by the fact that various operations defined by "surreal" recursions have a nice model theory. Examples: $(\mathbf{No}; \leqslant, +, \cdot)$ is a model of the theory of real closed fields; $(\mathbf{No}; \leqslant, +, \cdot, \exp)$ is a model of the theory of $\mathbb{R}_{\exp}$; and $(\mathbf{No}; \leqslant, +, \cdot, \partial_{\mathrm{BM}})$ is a model of the theory of H-closed fields. Each of these theories is model complete in a natural language. Is there a model theoretic reason that explains why this works so well?

Let us look at this in connection with the last example. Our aim is to define a derivation ∂ on $\mathbf{No}$ making it an H-field. Let $a \in \mathbf{No}$ be given for which we wish to define ∂a, and assume that ∂b has been defined for all $b \in L_a \cup R_a$. Let Δ_a be the class of all surreals b for which there exists a derivation ∂ on $\mathbf{No}$ with $\partial a = b$ and taking the prescribed values on $L_a \cup R_a$. Assembling all conditions that should be satisfied by ∂a, it is not hard to see that there exist sets $L, R \subseteq \mathbf{No}$ such that $\Delta_a = \{b \in \mathbf{No} : L < b < R\}$. We are left with two main questions: *When do we have $L < R$, thereby allowing us to define $\partial a = \{L \mid R\}$? Does this lead to a global definition of ∂ on* $\mathbf{No}$ *making it an H-closed field?* It might be of interest to isolate reasonable model theoretic conditions that imply the success of this type of construction. If the above construction does work, yet another question is whether the resulting derivation coincides with ∂_{BM}.

References

Norman L. Alling (1987). *Foundations of analysis over surreal number fields*. Vol. 141. North-Holland Mathematics Studies. Notas de Matemática [Mathematical Notes], 117. North-Holland Publishing Co., Amsterdam, pp. xvi+373. MR: 886475 (cit. on p. 20).

Emil Artin and Otto Schreier (1927). "Algebraische Konstruktion reeller Körper". *Abh. Math. Sem. Univ. Hamburg* 5.1, pp. 85–99. MR: 3069467 (cit. on p. 23).

Matthias Aschenbrenner, Lou van den Dries, and Joris van der Hoeven (2013). "Toward a model theory for transseries". *Notre Dame J. Form. Log.* 54.3-4, pp. 279–310. MR: 3091660 (cit. on p. 27).

– (Dec. 2015). "The surreal numbers as a universal H-field". J. Eur. Math. Soc. (JEMS), to appear. arXiv: 1512.02267 (cit. on pp. 34, 36).
– (2017a). *Asymptotic differential algebra and model theory of transseries*. Vol. 195. Ann. of Math. Stud. Princeton University Press, Princeton, NJ (cit. on pp. 20, 21, 25–30, 36).
– (2017b). "Dimension in the realm of transseries". In: *Ordered algebraic structures and related topics*. Vol. 697. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 23–39. MR: 3716064 (cit. on p. 29).
– (Jan. 2017c). "Maximal immediate extensions of valued differential fields". Proc. London Math. Soc., to appear. arXiv: 1701.06691 (cit. on p. 32).
Alessandro Berarducci and Vincenzo Mantova (Mar. 2017). "Transseries as germs of surreal functions". Trans. Amer. Math. Soc., to appear. arXiv: 1703.01995 (cit. on p. 38).
– (2018). "Surreal numbers, derivations and transseries". *J. Eur. Math. Soc. (JEMS)* 20.2, pp. 339–390. arXiv: 1503.00315. MR: 3760298 (cit. on pp. 36, 38).
Paul du Bois-Reymond (1871). "Sur la grandeur relative des infinis des fonctions". *Ann. Mat. Pura Appl.* 4, pp. 338–353 (cit. on p. 19).
– (1872). "Théorème général concernant la grandeur relative des infinis des fonctions et de leurs dérivées". *J. Reine Angew. Math.* 74, pp. 294–304. MR: 1579548 (cit. on p. 19).
– (1873). "Eine neue Theorie der Convergenz und Divergenz von Reihen mit positiven Gliedern". *J. Reine Angew. Math.* 76, pp. 61–91. MR: 1579577 (cit. on p. 19).
– (1875). "Ueber asymptotische Werthe, infinitäre Approximationen und infinitäre Auflösung von Gleichungen". *Math. Ann.* 8.3, pp. 363–414. MR: 1509850 (cit. on p. 19).
– (1877). "Ueber die Paradoxen des Infinitärcalcüls". *Math. Ann.* 11.2, pp. 149–167. MR: 1509910 (cit. on p. 19).
– (1882). *Die allgemeine Functionentheorie*. Verlag der H. Laupp'schen Buchhandlung, Tübingen (cit. on p. 19).
Émile Borel (1895). "Sur quelques points de la théorie des fonctions". *Ann. Sci. École Norm. Sup. (3)* 12, pp. 9–55. MR: 1508908 (cit. on p. 30).
– (1899). "Mémoire sur les séries divergentes". *Ann. Sci. École Norm. Sup. (3)* 16, pp. 9–131. MR: 1508965 (cit. on pp. 24, 32).
Michael Boshernitzan (1981). "An extension of Hardy's class L of 'orders of infinity'". *J. Analyse Math.* 39, pp. 235–255. MR: 632463 (cit. on p. 19).
– (1982). "New 'orders of infinity'". *J. Analyse Math.* 41, pp. 130–167. MR: 687948 (cit. on pp. 19, 29).
– (1986). "Hardy fields and existence of transexponential functions". *Aequationes Math.* 30.2-3, pp. 258–280. MR: 843667 (cit. on pp. 19, 31).
– (1987). "Second order differential equations over Hardy fields". *J. London Math. Soc. (2)* 35.1, pp. 109–120. MR: 871769 (cit. on pp. 19, 24, 29).

Nicolas Bourbaki (1951). *Élements de mathématique. XII. Premierère partie: Les structures fondamentales de l'analyse. Livre IV: Fonctions d'une variable réelle. (Théorie élémentaire). Chapitre IV: Équations différentielles. Chapitre V: Étude locale des fonctions. Chapitre VI: Développements tayloriens généralisés; formule sommatoire d'Euler-Maclaurin. Chapitre VII: La fonction gamma*. Actualités Sci. Ind., no. 1132. Hermann et Cie., Paris, pp. ii+200. MR: 0045774 (cit. on pp. 19, 24).

John H. Conway (1976). *On numbers and games*. London Mathematical Society Monographs, No. 6. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, pp. ix+238. MR: 0450066 (cit. on p. 19).

– (1994). "The surreals and the reals". In: *Real numbers, generalizations of the reals, and theories of continua*. Vol. 242. Synthese Lib. Kluwer Acad. Publ., Dordrecht, pp. 93–103. MR: 1340456 (cit. on p. 20).

Ovidiu Costin (2009). *Asymptotics and Borel summability*. Vol. 141. Chapman & Hall, CRC Monographs and Surveys in Pure and Applied Mathematics. CRC Press, Boca Raton, FL, pp. xiv+250. MR: 2474083 (cit. on p. 32).

Ovidiu Costin, Philip Ehrlich, and Harvey M. Friedman (May 2015). "Integration on the surreals: a conjecture of Conway, Kruskal and Norton". arXiv: 1505.02478 (cit. on p. 20).

Bernd I. Dahn and Peter Göring (1987). "Notes on exponential-logarithmic terms". *Fund. Math.* 127.1, pp. 45–50. MR: 883151 (cit. on p. 20).

Lou van den Dries (2000). "An intermediate value property for first-order differential polynomials". In: *Connections between model theory and algebraic and analytic geometry*. Vol. 6. Quad. Mat. Dept. Math., Seconda Univ. Napoli, Caserta, pp. 95–105. MR: 1930683 (cit. on p. 24).

Lou van den Dries and Philip Ehrlich (2001a). "Erratum to: 'Fields of surreal numbers and exponentiation'". *Fund. Math.* 168.3, pp. 295–297. MR: 1853411 (cit. on p. 20).

– (2001b). "Fields of surreal numbers and exponentiation". *Fund. Math.* 167.2, pp. 173–188. MR: 1816044 (cit. on p. 20).

Lou van den Dries, Angus Macintyre, and David Marker (1997). "Logarithmic-exponential power series". *J. London Math. Soc. (2)* 56.3, pp. 417–434. MR: 1610431 (cit. on pp. 20, 25, 26).

– (2001). "Logarithmic-exponential series". *Ann. Pure Appl. Logic* 111.1-2, pp. 61–113. MR: 1848569 (cit. on pp. 20, 26).

Jean Écalle (1992). *Introduction aux fonctions analysables et preuve constructive de la conjecture de Dulac*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, pp. ii+340. MR: 1399559 (cit. on pp. 20, 21, 27, 31, 32, 34).

G. A. Edgar (2010). "Transseries for beginners". *Real Anal. Exchange* 35.2, pp. 253–309. MR: 2683600 (cit. on p. 26).

Philip Ehrlich (2006). "The rise of non-Archimedean mathematics and the roots of a misconception. I. The emergence of non-Archimedean systems of magnitudes". *Arch. Hist. Exact Sci.* 60.1, pp. 1–121. MR: 2206281 (cit. on p. 19).

– (2012). "The absolute arithmetic continuum and the unification of all numbers great and small". *Bull. Symb. Log.* 18, pp. 1–45 (cit. on pp. 23, 24).

Paul Erdős, Leonard Gillman, and Melvin Henriksen (1955). "An isomorphism theorem for real-closed fields". *Ann. of Math. (2)* 61, pp. 542–554. MR: 0069161 (cit. on p. 23).

Jean Esterle (1977). "Solution d'un problème d'Erdős, Gillman et Henriksen et application à l'étude des homomorphismes de $\mathcal{C}(K)$". *Acta Math. Acad. Sci. Hungar.* 30.1-2, pp. 113–127. MR: 0482216 (cit. on p. 23).

Ulrich Felgner (2002). "Die Hausdorffsche Theorie der η_α-Mengen und ihre Wirkungsgeschichte". In: *Felix Hausdorff—gesammelte Werke*. Vol. II. Springer-Verlag, Berlin, pp. 645–674. MR: 2229142 (cit. on p. 19).

Gordon Fisher (1981). "The infinite and infinitesimal quantities of du Bois-Reymond and their reception". *Arch. Hist. Exact Sci.* 24.2, pp. 101–163. MR: 632567 (cit. on p. 19).

Harry Gonshor (1986). *An introduction to the theory of surreal numbers*. Vol. 110. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, pp. vi+192. MR: 872856 (cit. on pp. 20, 35).

Hans Hahn (1907). "Über die nichtarchimedischen Größensysteme". *S.-B. Akad. Wiss. Wien, Math.-naturw. Kl. Abt. IIa* 116, pp. 601–655 (cit. on p. 20).

Godfrey H. Hardy (1910). *Orders of infinity. The Infinitärcalcül of Paul du Bois-Reymond*. Vol. 12. Cambridge Tracts in Mathematics and Mathematical Physics. Cambridge Univ. Press, Cambridge, pp. v+62. MR: 0349922 (cit. on pp. 19, 25).

– (1912a). "Properties of logarithmico-exponential functions". *Proc. London Math. Soc. (2)* 10, pp. 54–90. MR: 1576038 (cit. on pp. 19, 24, 25).

– (1912b). "Some results concerning the behaviour at infinity of a real and continuous solution of an algebraic differential equation of the first order". *Proc. London Math. Soc. (2)* 10, pp. 451–468. MR: 1576033 (cit. on pp. 19, 24).

– (1913). "Oscillating Dirichlet's integrals: An essay in the 'Infinitärcalcül' of Paul du Bois-Reymond". *Quart. J. Pure Appl. Math* 44, 1–40 and 242–63 (cit. on p. 19).

Felix Hausdorff (1906a). "Untersuchungen über Ordnungstypen, I, II, III". *Sitzungsbericht der Königlich Sächsischen Gesellschaft für Wissenschaft Leipzig, Mathematische Physik* 58, pp. 106–169 (cit. on p. 19).

– (1906b). "Untersuchungen über Ordnungstypen, IV, V". *Sitzungsbericht der Königlich Sächsischen Gesellschaft für Wissenschaft Leipzig, Mathematische Physik* 59, pp. 84–159 (cit. on p. 19).

– (1909). "Die Graduierung nach dem Endverlauf". *Abh. Sächs. Akad. Wiss. Leipzig Math.-Natur. Kl.* 31, pp. 295–334 (cit. on pp. 19, 23).

Joris van der Hoeven (1997). *Asymptotique automatique*. PhD thesis. Université Paris VII, Paris, pp. vi+405. MR: 1458614 (cit. on pp. 20, 25, 34).

– (2001). "Operators on generalized power series". *Illinois J. Math.* 45.4, pp. 1161–1190. MR: 1894891 (cit. on p. 20).

– (2002). "A differential intermediate value theorem". In: *Differential equations and the Stokes phenomenon*. World Scientific, pp. 147–170 (cit. on p. 28).

– (2006). *Transseries and real differential algebra*. Vol. 1888. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. xii+255. MR: 2262194 (cit. on pp. 26, 27, 33, 34, 37).

– (2009). "Transserial Hardy fields". *Astérisque* 323, pp. 453–487 (cit. on p. 29).

Hellmuth Kneser (1949). "Reelle analytische Lösungen der Gleichung $\varphi(\varphi(x)) = e^x$ und verwandter Funktionalgleichungen". *J. Reine Angew. Math.* 187, pp. 56–67. MR: 0035385 (cit. on p. 31).

Donald E. Knuth (1974). *Surreal numbers*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, pp. iii+119. MR: 0472278 (cit. on p. 19).

Wolfgang Krull (1932). "Allgemeine Bewertungstheorie". *J. Reine Angew. Math.* 167, pp. 160–196. MR: 1581334 (cit. on p. 32).

Tullio Levi-Civita (1898). "Sui numeri transfiniti". *Atti Della R. Accademia Dei Lincei* 7, 91–96 and 113–121 (cit. on p. 20).

– (1892–93). "Sugli infiniti ed infinitesimi attuali quali elementi analitici". *Ist. Veneto Sci. Lett. Arti Atti Cl. Sci. Mat. Natur.* 4, pp. 1765–1815 (cit. on p. 20).

Vojislav Marić (1972). "Asymptotic behavior of solutions of a nonlinear differential equation of the first order". *J. Math. Anal. Appl.* 38, pp. 187–192. MR: 0306640 (cit. on p. 24).

Frédéric Menous (1999). "Les bonnes moyennes uniformisantes et une application à la resommation réelle". *Ann. Fac. Sci. Toulouse Math. (6)* 8.4, pp. 579–628. MR: 1815157 (cit. on p. 32).

Chris Miller (2012). "Basics of o-minimality and Hardy fields". In: *Lecture notes on o-minimal structures and real analytic geometry*. Vol. 62. Fields Inst. Commun. Springer, New York, pp. 43–69. MR: 2976990 (cit. on pp. 19, 31).

Abraham Robinson (1972). "On the real closure of a Hardy field". In: *Theory of sets and topology (in honour of Felix Hausdorff, 1868–1942)*. VEB Deutsch. Verlag Wissensch., Berlin, pp. 427–433. MR: 0340225 (cit. on p. 24).

Maxwell Rosenlicht (1983a). "Hardy fields". *J. Math. Anal. Appl.* 93.2, pp. 297–311. MR: 700146 (cit. on pp. 19, 24).

– (1983b). "The rank of a Hardy field". *Trans. Amer. Math. Soc.* 280.2, pp. 659–671. MR: 716843 (cit. on p. 19).

– (1984). "Rank change on adjoining real powers to Hardy fields". *Trans. Amer. Math. Soc.* 284.2, pp. 829–836. MR: 743747 (cit. on p. 19).

– (1987). "Growth properties of functions in Hardy fields". *Trans. Amer. Math. Soc.* 299.1, pp. 261–272. MR: 869411 (cit. on p. 19).
– (1995). "Asymptotic solutions of $Y'' = F(x)Y$". *J. Math. Anal. Appl.* 189.3, pp. 640–650. MR: 1312544 (cit. on p. 19).
Simon Rubinstein-Salzedo and Ashvin Swaminathan (2014). "Analysis on surreal numbers". *J. Log. Anal.* 6, Paper 5, 39. MR: 3267278 (cit. on p. 20).
Michael Ch. Schmeling (2001). "Corps de transséries". PhD thesis. Université Paris VII (cit. on pp. 20, 25, 26, 33–35).
Michael F. Singer (1975). "Asymptotic behavior of solutions of differential equations and Hardy fields: preliminary report". Manuscript (cit. on p. 24).
Gunnar Sjödin (1971). "Hardy-fields". *Ark. Mat.* 8, pp. 217–237. MR: 0294577 (cit. on pp. 24, 30, 31).
Alfred Tarski (1951). *A decision method for elementary algebra and geometry*. 2nd ed. University of California Press, Berkeley and Los Angeles, Calif., pp. iii+63. MR: 0044472 (cit. on p. 20).
Alex J. Wilkie (1996). "Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function". *J. Amer. Math. Soc.* 9.4, pp. 1051–1094. MR: 1398816 (cit. on pp. 20, 22).

Received 2017-11-17.

MATTHIAS ASCHENBRENNER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CA 90095
USA
matthias@math.ucla.edu

LOU VAN DEN DRIES
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, IL 61801
USA
vddries@math.uiuc.edu

JORIS VAN DER HOEVEN
ÉCOLE POLYTECHNIQUE
91128 PALAISEAU CEDEX
FRANCE
vdhoeven@lix.polytechnique.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (43–62)

TOWARDS A THEORY OF DEFINABLE SETS

Stephen Jackson

Abstract

The subject of descriptive set theory is traditionally concerned with the theory of definable subsets of Polish spaces. By introducing large cardinal/determinacy axioms, a theory of definable subsets of Polish spaces and their associated ordinals has been developed over the last several decades which extends far up in the definability hierarchy. Recently, much interest has been focused on trying to extend the theory of definable objects to more general types of sets, not necessarily subsets of a Polish space or an ordinal. A large class of these objects are represented by equivalence relations on Polish spaces. Even for some of the simpler of these relations, an interesting combinatorial theory is emerging. We consider both problems of extending further the theory of definable subsets of Polish spaces, and that of determining the structure of these new types of definable sets.

1 Introduction and background

The field of descriptive set theory traditionally in concerned with the theory of definable sets in Polish spaces (complete, separable, metric spaces). As all uncountable Polish spaces are isomorphic by a Borel function, it is customary to refer to the elements of any of several standard Polish spaces as “reals.” Aside from $\mathbb{R}$, familiar examples include the Baire space ω^ω (homeomorphic to the space of irrationals in $\mathbb{R}$; here ω denotes the set of natural numbers), and 2^ω (homeomorphic to the Cantor set in $\mathbb{R}$; here $2 = \{0, 1\}$). In the latter two cases, ω is endowed with the discrete topology, and ω^ω or 2^ω with the product topology. Note that if G is any countable discrete group, then 2^G is likewise homeomorphic to the Cantor set, so it is naturally a compact Polish space.

Using the axiom of choice, AC, “pathological” sets with a variety of properties can be constructed. Examples include Vitali sets (non-measurable sets), Bernstein sets (a set such that neither the set nor its complement contains a closed uncountable set), Lusin sets (a set of reals which meets every meager set in a countable set), and Sierpinski sets (a set

MSC2010: primary 03E15; secondary 03E60.

which meets every measure 0 set in a countable set). A theme of descriptive set theory is that if we restrict out attention to "definable" sets, then these pathologies disappear and a reasonable structure theory emerges. The notion of definable is made precise through hierarchies of collections of sets of increasing complexity. A *pointclass* $\boldsymbol{\Gamma}$ is a collection of subsets of Polish spaces which is closed under inverse images by continuous functions, that is, if $f: X \to Y$ is continuous and $A \subseteq Y$ is in $\boldsymbol{\Gamma}$, then $f^{-1}(A)$ is also in $\boldsymbol{\Gamma}$. A basic example is the pointclass of Borel sets, the smallest collection containing the open and closed sets and closed under countable unions and intersections. The Borel sets are stratified into the *Borel hierarchy*, the pointclasses $\boldsymbol{\Sigma}^0_\alpha$, $\boldsymbol{\Pi}^0_\alpha$, and $\boldsymbol{\Delta}^0_\alpha$, for $\alpha < \omega_1$. Here $\boldsymbol{\Sigma}^0_1 \upharpoonright X$ is the collection of open sets in the Polish space X, $\boldsymbol{\Pi}^0_1 \upharpoonright X$ the closed sets in X, $\boldsymbol{\Delta}^0_1 \upharpoonright X = \boldsymbol{\Sigma}^0_1 \upharpoonright X \cap \boldsymbol{\Pi}^0_1 \upharpoonright X$, and in general $A \in \boldsymbol{\Sigma}^0_\alpha$ if $A = \bigcup_n A_n$ where each $A_n \in \boldsymbol{\Pi}^0_{\beta_n}$ for some $\beta_n < \alpha$. Likewise, $A \in \boldsymbol{\Pi}^0_\alpha$ if $A = \bigcap_n A_n$ with each $A_n \in \boldsymbol{\Sigma}^0_{\beta_n}$ for some $\beta_n < \alpha$. Also, we define $\boldsymbol{\Delta}^0_\alpha = \boldsymbol{\Sigma}^0_\alpha \cap \boldsymbol{\Pi}^0_\alpha$. It is a classical fact that the Borel sets in any Polish space have the perfect set property (if they are uncountable then they contain a perfect set, or equivalently an uncountable closed set), and are Lebesgue measurable and have the Baire property. Thus, they cannot be any of the above types of pathological sets. Another example of a "regularity property" for sets is the *Ramsey property* for the set $A \subseteq [\omega]^\omega$ (here $[H]^\omega$ denotes the set of infinite subsets of H, which we can identify with the set of increasing functions from ω to ω) which asserts that there is an infinite set $H \subseteq \omega$ such that either $[H]^\omega \subseteq A$ or $[H]^\omega \subseteq \omega^\omega - A$. Again, all Borel sets have this regularity property, this being a theorem of Galvin and Prikry (in fact the Borel sets are *completely Ramsey*, a somewhat stronger version of the Ramsey property).

The hierarchy of definable sets extends far beyond the Borel sets. The next hierarchy after the Borel sets is the *projective hierarchy*, so called because the main operation used in generating the hierarchy is projection from a product $X \times Y$ of Polish spaces to X. The *analytic*, or $\boldsymbol{\Sigma}^1_1$ sets, are defined by projecting closed (or equivalently Borel) sets: $A \subseteq X$ is $\boldsymbol{\Sigma}^1_1$ iff there is a closed set $F \subseteq X \times \omega^\omega$ such that $x \in A$ iff $\exists y\ (x, y) \in F$. In more succinct notation, we write $\boldsymbol{\Sigma}^1_1 \upharpoonright X = \exists^{\omega^\omega} \boldsymbol{\Pi}^0_1 \upharpoonright (X \times \omega^\omega)$, where $\exists^Y$ denotes the operation of applying existential quantification over Y. A set $A \subseteq X$ is *co-analytic*, or $\boldsymbol{\Pi}^1_1$, if it is the complement of an analytic set. That is, $\boldsymbol{\Pi}^1_1$ is the *dual pointclass* of $\boldsymbol{\Sigma}^1_1$, which we write as $\boldsymbol{\Pi}^1_1 = \check{\boldsymbol{\Sigma}}^1_1$. We set $\boldsymbol{\Delta}^1_1 = \boldsymbol{\Sigma}^1_1 \cap \boldsymbol{\Pi}^1_1$ to be the sets that are both $\boldsymbol{\Sigma}^1_1$ and $\boldsymbol{\Pi}^1_1$. A classical theorem of Suslin states that $\boldsymbol{\Delta}^1_1$ is the collection of Borel sets, so the projective hierarchy begins where the Borel hierarchy ends. We continue to define the $\boldsymbol{\Sigma}^1_n$, $\boldsymbol{\Pi}^1_n$, and $\boldsymbol{\Delta}^1_n$ sets for all $n \in \omega$ by setting $\boldsymbol{\Sigma}^1_{n+1} = \exists^{\omega^\omega} \boldsymbol{\Pi}^1_n$, $\boldsymbol{\Pi}^1_{n+1} = \check{\boldsymbol{\Sigma}}^1_{n+1}$ (or equivalently, $\boldsymbol{\Pi}^1_{n+1} = \forall^{\omega^\omega} \boldsymbol{\Sigma}^1_n$), and $\boldsymbol{\Delta}^1_{n+1} = \boldsymbol{\Sigma}^1_{n+1} \cap \boldsymbol{\Pi}^1_{n+1}$. The projective hierarchy is important because it includes all of the sets of conventional analysis. In fact, the sets of analysis generally occur at the first or second levels of this hierarchy. In any uncountable Polish space, all of the levels of the Borel and projective hierarchies are distinct, that is,

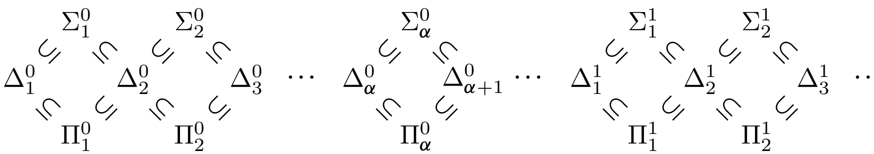


Figure 1: The Borel and projective hierarchies.

there is no collapsing in either of these hierarchies. The inclusions of pointclasses within these hierarchies is shown in Figure 1.

Beginning with the fundamental work of Gödel, it was realized that there were strong limits to how much further far one could extend the regularity results for Borel sets working just in ZFC set theory (the set theory of "ordinary mathematics"). For example, the $\boldsymbol{\Sigma}^1_1$ sets have the perfect set property, but it is consistent with ZFC that the $\boldsymbol{\Pi}^1_1$ sets do not. Likewise, while the $\boldsymbol{\Sigma}^1_1$ and $\boldsymbol{\Pi}^1_1$ sets are all Lebesgue measurable and have the Baire property, it is consistent with ZFC that the collection of $\boldsymbol{\Delta}^1_2$ sets does not. A theorem of Silver asserts that the $\boldsymbol{\Sigma}^1_1$ and $\boldsymbol{\Pi}^1_1$ sets are all (completely) Ramsey, but it is again consistent that there are $\boldsymbol{\Delta}^1_2$ sets which are not. Thus, in order to extend the theory further, one must assume additional axioms which go beyond the ZFC axioms. There are currently two main axiom schemes for doing this: large cardinal axioms and determinacy axioms. Large cardinal axioms, which are generally meant to be added to the ZFC axioms, assert that cardinals κ with certain properties exist which cannot be shown to exist just from ZFC. Determinacy axioms, on the other hand, assert that certain two-player games are determined. If $A \subseteq \omega^\omega$, then we associate a two-player integer game $G(A)$ to A in a natural way: the players I and II alternate picking integers $x(n) \in \omega$ as shown in Figure 2. They thereby jointly build an $x = (x(0), x(1), \dots) \in \omega^\omega$. Player I wins the game iff $x \in A$.

	I	$x(0)$		$x(2)$			
$G(A)$						$\cdots$	$x \in \omega^\omega$
	II		$x(1)$		$x(3)$		

Figure 2: The basic game $G(A)$.

The notions of a winning strategy for one of the players, and of the game being determined (i.e., one of the players has a winning strategy) are defined in the natural manner. The axiom of determinacy, AD, is the assertion that $G(A)$ is determined for all $A \subseteq \omega^\omega$.

This axiom contradicts the axiom of choice, AC, but is meant to be an axiom for certain inner models of the full universe of sets V for which the sets are, in some sense, definable. If $\mathbf{\Gamma}$ is a pointclass, then $\det(\mathbf{\Gamma})$ is the statement that $G(A)$ is determined for all $A \subseteq \omega^\omega$ in $\mathbf{\Gamma}$. A celebrated theorem of Martin says all Borel games are determined. More generally, if X is any set and $A \subseteq X^\omega$ is Borel in the product topology on X^ω, where X is given the discrete topology, then (in ZFC) the game $G(A)$ is determined. In fact, a version of this result holds in ZF (without choice), where now every every Borel game $G(A)$ for $A \subseteq X^\omega$ is *quasi-determined* (see Moschovakis [1980]). Here a quasi-strategy is like a strategy except it is multi-valued. Results of Martin and Harrington show that $\mathsf{ZFC} + \det(\mathbf{\Sigma}^1_1)$ is strictly stronger than ZFC. The assertions $\det(\mathbf{\Sigma}^1_n)$ are strictly increasing in strength, and projective determinacy, PD, is the statement that all projective games are determined. The model $L(\mathbb{R})$ is the smallest model of set theory containing all of the reals and ordinals. Every set of reals in this model is definable in this model by a formula using only ordinal and real parameters. Because of this, the axiom $\mathsf{AD}^{L(\mathbb{R})}$ that all $A \in L(\mathbb{R})$ are determined is a plausible axiom. Since strategies in integer games are essentially reals, if $\mathsf{AD}^{L(\mathbb{R})}$ holds then in fact $L(\mathbb{R})$ satisfies the axiom AD. The model $L(\mathbb{R})$ is thus the smallest candidate inner model (containing all the ordinals) which satisfies AD. More generally, if M is any inner model satisfying AD, then we may consider the sets in M as being, in some abstract sense, definable.

Work of Martin, Steel, and Woodin in the 80's established the precise connection between determinacy axioms and large cardinal axioms. It was shown, for example, that $\mathsf{ZFC} + \exists n$ Woodin cardinals $+\exists$ a larger measurable cardinal implies $\det(\mathbf{\Sigma}^1_{n+1})$. Also, AD is equiconsistent with the existence of ω many Woodin cardinals, and $\mathsf{AD}^{L(\mathbb{R})}$ is implied by the existence of ω many Woodin cardinals and a measurable cardinal above them. These fundamental results lend respectability to the determinacy axioms and show that not only is $\mathsf{AD}^{L(\mathbb{R})}$ an intuitively appealing axiom, but that it is actually implied by large cardinal axioms out in the ZFC universe.

Moving past AD, we many consider the axiom AD_X which asserts that every game played on the set X, that is, where $A \subseteq X^\omega$, is determined. This axiom is inconsistent for $X = \omega_1$ or $X = \mathcal{P}(\mathbb{R})$, but for $X = \mathbb{R}$ the axiom of real game determinacy $\mathsf{AD}_{\mathbb{R}}$ is reasonable. This axiom is significantly stronger than AD, and cannot hold in the minimal model $L(\mathbb{R})$ of AD, or in any model of the form $L(T, \mathbb{R})$, for $T \subseteq \mathrm{On}$. Woodin has identified the exact consistency strength of $\mathsf{AD}_{\mathbb{R}}$ in the large cardinal hierarchy as well. Thus we have a progression of the determinacy axioms starting with $\det(\mathbf{\Delta}^1_1)$, which is a theorem of ZFC, to $\det(\mathbf{\Sigma}^1_1)$, $\det(\mathbf{\Sigma}^1_n)$, PD, AD, and $\mathsf{AD}_{\mathbb{R}}$. There are also stronger determinacy axioms than $\mathsf{AD}_{\mathbb{R}}$ involving "long games," but we will not need to consider these here. We note that in any model of AD the sets of reals fall into a single hierarchy, called the Wadge hierarchy, which gives a far-reaching generalization of the Borel and projective hierarchies. In particular, in these models we can define the $\boldsymbol{\Sigma}^1_\alpha$, $\boldsymbol{\Pi}^1_\alpha$ classes for all $\alpha < \Theta$, where Θ is the length of the Wadge hierarchy in the model. Thus, these higher level analogs of the projective sets are defined and extend throughout the entire Wadge hierarchy of sets of reals.

Beginning in the 60's, and continuing to the present, it was realized that determinacy axioms were a powerful tool which allowed the classical results for Borel and analytic sets to be extended to larger classes of sets. Work of Kechris, Martin, Moschovakis, Solovay, Steel, Woodin, and others showed that assuming determinacy axioms, and in particular assuming AD, one could propagate a structural theory similar to the ZFC theory of Borel and analytic sets. This theory is largely presented in terms of *scales* and *Suslin cardinals*, and gives a tight connection between the theory of the sets of reals in a pointclass $\boldsymbol{\Gamma}$ and the properties of an ordinal $\delta(\boldsymbol{\Gamma})$ associated to the pointclass. The notion of a scale was isolated by Moschovakis, and has origins in the Novikov-Kondo proof of $\boldsymbol{\Pi}^1_1$ uniformization. We recall the following definition. By a tree T on a set X we mean a $T \subseteq X^{<\omega}$ which is closed under subsequence, that is, if $s \in T$ and m is less than the length of s, then $s \restriction m \in T$. We let $[T] = \{x \in X^\omega : \forall n\ x \restriction n \in T\}$ be the set of infinite branches (or body) of T.

Definition 1. We say a set $A \subseteq \omega^\omega$ is κ-Suslin, for $\kappa \in \mathrm{On}$, if there is a tree $T \subseteq (\omega \times \kappa)^{<\omega}$ such that $A = p[T] = \{x \in \omega^\omega : \exists f \in \kappa^\omega\ (x, f) \in [T]\}$.

We say κ is a Suslin cardinal if there is a set A which is κ-Suslin but not λ-Suslin for any $\lambda < \kappa$. The notions of semi-scale and scale are a more algebraic reformulation of having Suslin representations, presented in terms of norms $\varphi_n : A \to \kappa$. In fact, being κ-Suslin is equivalent to having a semi-scale with norms to κ, and also equivalent to having a scale with norms to κ. We refer the reader to Moschovakis [ibid.] for the precise definitions of semi-scales, scales, and the scale property for a pointclass.

Assuming AD, we can propagate the scale/Suslin cardinal analysis past the $\boldsymbol{\Sigma}^1_1$, $\boldsymbol{\Pi}^1_1$ levels to the entire projective hierarchy and beyond. In Jackson [2010] one can find a presentation of the complete scale and Suslin cardinal analysis from AD. This analysis, though it extends throughout the full extent of the Suslin cardinals, presents the theory in terms of the ordinals $\delta(\boldsymbol{\Gamma})$. A much more detailed inductive analysis is necessary to analyze these ordinals and describe the cardinal structure below them. In Jackson [1999] and Jackson [1988] this analysis is described through the projective hierarchy. The extent of this analysis is currently far short of the extent of scales, and so much about the general cardinal structure of determinacy models remains unknown.

To give one example of the consequences of this analysis, we first recall that it is classical fact (proved in ZFC) that every $\boldsymbol{\Sigma}^1_1$ or $\boldsymbol{\Pi}^1_1$ set is ω_2-Borel, that is, is in the smallest collection containing the open and closed sets and closed under unions and intersections of length $< \omega_2$. In fact, every $\boldsymbol{\Sigma}^1_1$ or $\boldsymbol{\Pi}^1_1$ set is an ω_1 union of Borel sets (a proof can be found in Moschovakis [1980], A. S. Kechris [1978], or Jackson [2010]). From the above mentioned inductive analysis we get the following extension of this result, assuming determinacy holds for the sets in $L(\mathbb{R})$ (see Jackson [1989]).

Theorem 2. *Assume* $\mathsf{ZFC} + \mathsf{AD}^{L(\mathbb{R})}$. *Then every projective set is* ω_ω*-Borel.*

Moving forward, in trying to develop the theory of definable objects from stronger set-theoretic axioms, there are two main directions to pursue. The first is to extend this theory of sets of reals and their associated ordinals further, and to to attempt to describe the entire cardinal structure of determinacy models. We might refer to this as extending the theory of "reals and ordinals." A second direction is to study more general types of objects, moving past those that be identified with sets in a Polish space or wellordered sets. Of course, the study of these more general definable objects encompasses the first direction, but the point is that we can advance the study of these more general objects without having the complete theory of the cardinals structure in hand.

In Section 2 we describe in a little more detail some of the progress in developing the theory of "reals and ordinals" and problems that are reasonably aligned with this program. We describe some of the recent progress various researchers have made, in particular using new techniques from inner model theory. This emerging area of "descriptive inner model theory" holds much promise for future progress in this area. We also mention some of the old questions and conjectures which are still around and which may serve as a benchmark for further progress. In Section 3 we consider some questions related to more general types of objects. Here we see an interesting and fascinating combinatorial structure beginning to emerge. The focus here is not so much on extending the theory to higher and higher pointclasses, but to understand how the new nature of these objects affects their combinatorial structure. Thus, we frequently consider problems at the Borel level, where the sets and functions used in the definitions of the objects are Borel, or even continuous/clopen. Recent years have seen a growing interest in this study of "Borel combinatorics" and its connections with other areas such as ergodic theory, geometric group theory, and descriptive set theory.

2 The theory of reals and ordinals

A well-known consequence of AD is that all sets of reals have the perfect set property, are measurable (with respect to any Borel measure), and have the Baire property. It follows that we have the Fubini theorem and its analog for category, the Kuratowski-Ulam theorem, for arbitrary sets $A \subseteq X \times Y$ in products of Polish spaces. We then also have full additivity of measure and category, that is, an arbitrary well-ordered union of meager (or measure 0) sets is meager (measure 0). In particular, from the perfect set property we have that there are only two possibilities for the cardinality of a set in reals in a determinacy model: countable and the size of the continuum. We note that one must be careful with the term "cardinality" in a model without AC as, for example a map from a set X onto a set Y does not necessarily yield a map from Y into X (in a model of AD there are maps from $\mathbb{R}$ onto any ordinal $\alpha < \Theta$, which is very large in the $\aleph_\beta$ hierarchy, but there is only an injection from α to $\mathbb{R}$ if α is countable). Nevertheless, if a set of reals contains a perfect set, then it is in bijection with $\mathbb{R}$.

The cardinal structure inside a model of determinacy is interesting and non-trivial. As we indicated before, the cardinal structure is closely connected with certain associated pointclasses. At the projective level, the ordinals associated to these classes are called the *projective ordinals*. More precisely, let

$$\delta^1_n = \delta(\mathbf{\Pi}^1_n) = \sup\{|\preceq| \colon \preceq \text{ is a prewellordering of } \omega^\omega \text{ in } \mathbf{\Delta}^1_n\}$$

where a prewellordering $\preceq$ is a reflexive, transitive, connected relation whose strict part $(x \prec y \leftrightarrow (x \preceq y) \wedge \neg(y \preceq x))$ is wellfounded and $|\preceq|$ denotes its length. The work of Kechris, Kunen, Martin, Moschovakis, and Solovay established the basic properties of the $\boldsymbol{\delta}^1_n$, and computed their values for $n \leq 4$ (these results can be found in A. S. Kechris [1978] of Jackson [2010]). The author computed their values for all n and described the structure of the cardinals below their supremeum (c.f. Jackson [1999] and Jackson [1988]). The Suslin cardinals below their supremum are the odd projective ordinals $\boldsymbol{\delta}^1_{2n+1}$ and their cardinal predecessors $\lambda^1_{2n+1} = (\boldsymbol{\delta}^1_{2n+1})^-$. The $\boldsymbol{\delta}^1_{2n+1}$-Suslin sets are the $\mathbf{\Sigma}^1_{2n+2}$ sets, and the λ^1_{2n+1}-Suslin sets are the $\mathbf{\Sigma}^1_{2n+1}$ sets. The cardinal structure below the supremum of the projective ordinals reveals some interesting and delicate patterns. The projective ordinals are all regular cardinals and the even ones are the successors of the odd ones, $\boldsymbol{\delta}^1_{2n+2} = (\boldsymbol{\delta}^1_{2n+1})^+$ (this was known from early work), and there are exactly $2^{n+1}-1$ many regular cardinals strictly between $\boldsymbol{\delta}^1_{2n+1}$ and $\boldsymbol{\delta}^1_{2n+3}$. The other cardinals between these two odd projective ordinals are all singular of cofinality one of these $2^{n+1} - 1$ cardinals. The values of the $\boldsymbol{\delta}^1_n$ can be computed exactly. The result is that $\boldsymbol{\delta}^1_{2n+1} = \omega_{e(2n-1)+1}$, where $e(0) = 1$ and $e(n+1) = \omega^{e(n)}$ (ordinal exponentiation). Also, the exact cofinalities of the cardinals below the projective ordinals can be computed (see Jackson and Khafizov [2016] and Jackson and Löwe [2013]). Figure 3 shows some of the cardinal structure below the projective ordinals. Note that the three regular cardinals between $\boldsymbol{\delta}^1_3$ and $\boldsymbol{\delta}^1_5$ are $\boldsymbol{\delta}^1_4 = \omega_{\omega+2}$, $\omega_{\omega\cdot 2+1}$, and $\omega_{\omega^\omega+1}$.

The detailed inductive analysis which provides the above mentioned analysis of cardinal cofinalities does not currently generalize to arbitrary levels of the Wadge hierarchy. While it does extend past the projective sets, likely to the first weakly inaccessible cardinal, it is known that the methods do not extend to the first "inductive like" pointclass (a non-selfdual pointclass closed under real quantification). Thus, some questions about cardinal cofinalities are still open past the projective hierarchy. To take an example, within

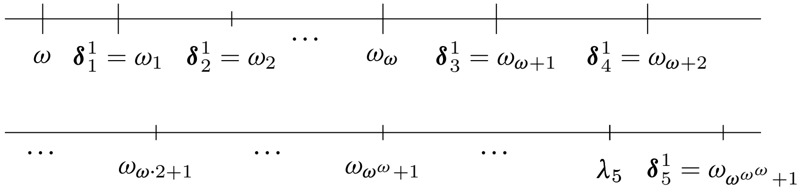


Figure 3: The cardinal structure below the projective ordinals.

the projective hierarchy there are never more than two regular cardinals in a row. Past the projective hierarchy, it is known that there can be three regular cardinals in a row (in Apter, Jackson, and B. Löwe [2013] this and stronger results are shown), however the evidence seems to suggest that there cannot be four in a row. This seems like a reasonable benchmark for understanding the cardinal structure, so we state:

Conjecture 3. Assuming AD, there does not exists a cardinal $\kappa < \Theta$ such that κ, κ^+, κ^{++}, and κ^{+++} are all regular.

Aside from the cofinalities of the cardinals, there are other interesting combinatorial properties of the cardinals which we may consider. One class of these concerns partition properties. In the Erdös-Rado partition notation we write $\kappa \to (\kappa)^\lambda$ if for all partitions $\mathcal{P} \colon \kappa^\lambda \to \{0,1\}$ of the increasing functions from λ to κ into two pieces, there is a set $H \subseteq \kappa$ of size κ and an $i \in \{0,1\}$ such that $\mathcal{P} \restriction H^\lambda = i$. The statement that all sets $A \subseteq \omega^\omega$ are Ramsey is the strong partition property for $\kappa = \omega$. This follows from AD^+, a technical strengthening of AD introduced by Woodin which has found many applications in determinacy theory (it is not known whether AD suffices for this result). Assuming AD, the cardinal analysis shows that all regular Suslin cardinals below the projective ordinals, which are just the $\boldsymbol{\delta}^1_{2n+1}$, have the strong partition property. This is also not known to extend to arbitrary levels, so we ask:

Conjecture 4. Assuming AD, every regular Suslin cardinal has the strong partition property.

It is shown in Jackson [2011] that for the (finitely many) regular cardinals κ between $\boldsymbol{\delta}^1_{2n+1}$ and $\boldsymbol{\delta}^1_{2n+3}$ we have $\kappa \to (\kappa)^{\boldsymbol{\delta}^1_{2n+1}}$, but $\kappa \not\to (\kappa)^{\boldsymbol{\delta}^1_{2n+2}}$. This leads to the following general problem.

Problem 5. Assume AD. Determine for each regular $\kappa < \Theta$ the λ such that $\kappa \to (\kappa)^\lambda$.

We close this section by results concerning three large cardinal notions interpreted in models of determinacy. Namely, we consider the notions of Jónsson cardinal, measurable cardinal, and supercompact cardinal. For $\kappa < \lambda < \theta$, we say that κ is λ-supercompact if there is a fine, normal measure ν on $\wp_\kappa(\lambda)$ (the subsets of λ of size less than κ). Here fine means that if $A \in \wp_\kappa(\lambda)$ then $\{B : B \supseteq A\}$ has ν measure one. Normal means that if $f : \wp_\kappa(\lambda) \to \lambda$ is such that $f(A) \in A$ for ν almost all A, then f is constant ν almost everywhere. Building on H. S. Becker and Jackson [2001] it is shown in Jackson [2001] that assuming AD^+ that every regular cardinal κ which is either a Suslin cardinal or the successor of a Suslin cardinal is λ-supercompact for all $\lambda < \Theta$. Again, this leads to a general problem:

Problem 6. Assume AD. For which regular κ and $\lambda > \kappa$ is κ λ-supercompact?

We say a cardinal κ is measurable if there is a non-principal, κ-complete ultrafilter μ of κ. We recall that assuming AD, every ultrafilter on a set is countably additive, that is, is a measure. It is not difficult to show that the regular Suslin cardinals are measurable, but for general regular cardinals the problem seems to require new methods. Specifically, methods of inner model theory have begun to play an important role in determinacy theory. In Steel [1995] Steel made an important breakthrough by using progress in inner model theory to analyze the inner model $\mathrm{HOD}^{L(\mathbb{R})}$ of $L(\mathbb{R})$ assuming a large cardinal/determinacy hypothesis, which Woodin improved to just assuming $\mathsf{AD}^{L(\mathbb{R})}$. A consequence of this analysis is:

Theorem 7 (Steel)**.** *Assume $\mathsf{AD} + V = L(\mathbb{R})$. Then every regular cardinal $\kappa < \Theta$ is measurable.*

As part of the HOD analysis, it is also shown that $\mathrm{HOD}^{L(\mathbb{R})}$ satisfies the GCH. Although this is a result about the model HOD, by "relativizing" it (i.e., using the fact that every set in $L(\mathbb{R})$ is definable from an ordinal and a real) we get:

Theorem 8 (Steel)**.** *Assume $\mathsf{AD} + V = L(\mathbb{R})$. Then for any $\kappa < \Theta$, any wellordered sequence of subsets of κ has length $< \kappa^+$.*

Again, the previous result was known to hold previously for Suslin cardinals κ. It is not known how to obtain either of the two previous theorems by direct determinacy arguments.

In a similar vein one can use the HOD analysis to prove a result concerning Jónsson cardinals. A cardinal κ is said to be Jonsson if for all $f : \kappa^{<\omega} \to \kappa$ there an $A \subseteq \kappa$ with $|A| = \kappa$ such that $f(A^{<\omega}) \neq \kappa$. Thus, the Jonsson property is a weak form of the partition property, weaker than being measurable. From the HOD analysis we get Jackson, Ketchersid, Schlutzenberg, and Woodin [2014]:

Theorem 9 (J, Ketchersid, Schlutzenberg, Woodin)**.** *Assume $\mathsf{AD}^{L(\mathbb{R})}$. Then every cardinal $\kappa < \Theta$ is Jónsson.*

Using additional arguments, Woodin has extended the last two theorems to models of AD^+. It remains open if these techniques can be extended to answer Conjectures 3, 4.

3 More general sets

In Section 2 we considered the problem of developing the theory of definable sets of reals and ordinals. The theory at the lower levels of the definability hierarchy seems fairly well established, though many interesting problems remain in extending this theory to higher levels. As we described, this theory is developed assuming stronger axioms than ZFC. In this section we consider the problem of developing the theory of more general types of sets. To motivate some of the basic objects of study, consider the model $L(\mathbb{R})$. In this model, every set $A \subseteq L_\Theta(\mathbb{R})$ is the surjective image of $\mathbb{R}$. Say $\pi : \mathbb{R} \to A$ is an onto map. This defines naturally an equivalence relation E on ω^ω, namely, $x\ E\ y$ iff $\pi(x) = \pi(y)$. It follows that A is in bijection with the quotient space $\mathbb{R}/\sim$. So, all of the sets in this model, at least those of rank $< \Theta$, can be identified with the set of equivalence classes (quotient space) of an equivalence relation on the Polish space $\mathbb{R}$. Thus, the collection of sets which can be represented as quotient spaces by equivalence relations on Polish spaces is a quite large collection, greatly extending the collection of sets which can be identified with a subset of a Polish space, or which can be wellordered (identified with an ordinal).

The simplest equivalence relations on Polish spaces are the *smooth* ones. We say (X, E) is smooth if there is a Borel map $f : X \to Y$, for some Polish space Y such that $x\ E\ y$ iff $f(x) = f(y)$. We can, of course, replace "Borel" with more liberal notions of definable, but in most cases this is a good stand-in for the more general case. In this case, the quotient space X/E can be identified with a subset of Y, namely the range of f. Conversely, any subset A of a Polish space X can be identified with the quotient space of a smooth equivalence relation on X. So, if (X, E) is smooth, or if the classes can be wellordered, then we are in the case of Section 2. So, from the point of view of introducing new types of definable objects, we consider these to be "trivial" equivalence relations.

One of the simplest non-trivial equivalence relations is the equivalence relation of eventual agreement on 2^ω known as E_0: $x\ E_0\ y$ iff $\exists n\ \forall m \geq n\ (x(m) = y(m))$. Note that the E_0 relation is a simple ($\mathbf{\Sigma}^0_2$) Borel equivalence relation. There is a natural action of the group $\mathbb{Z}$ on 2^ω called the odometer action which is defined by $1 \cdot x$ (1 being the generator $\mathbb{Z}$) is obtained by adding 1 to x viewed as an infinite binary expansion (with $x(0)$ being the least significant digit). This $\mathbb{Z}$ action is defined on all classes except $[\bar{0}]$ and $[\bar{1}]$, which are the constant 0 and 1 reals. The natural definition of the odometer map on the classes $[\bar{0}]$ and $[\bar{1}]$ amalgamates these two classes, but we can redefine the map on these two classes so that the $\mathbb{Z}$ action generates the E_0 equivalence relation, that is, $x\ E_0\ y$ iff $\exists n \in \mathbb{Z}\ (n \cdot x = y)$. The natural Bernoulli measure on 2^ω is invariant under this action (the redefinition on the two distinguished classes doesn't affect anything). It follows that there cannot be a Borel, or even measurable, *selector* for E_0, that is, a set S which meets every E_0 class in exactly one point. So, E_0 is not smooth. Of course, with AC one can form a selector by simply picking an element for each E_0 class, but this does not result in a definable set. This simple argument is just the standard Vitali argument for the construction of (non-definable) non-measurable set. From our current point of view, focusing on definable objects, the quotient space of E_0 is a new type of object, not given by a subset of a Polish space or an ordinal. This immediately raises a general question: what can we say about the structure of these definable equivalence relations on Polish spaces? As the above example shows, restricting the notion of definable to Borel still captures the main essence of the new phenomenon, and thus we led to the study of Borel equivalence relations on Polish spaces.

The motivation expressed in the above arguments for studying Borel equivalence relations is only one of many such possible. For example, classical dynamic can be viewed as the study of Borel actions of the group $\mathbb{Z}$ on Polish spaces, frequently equipped with other structure such an an invariant probability measure on X. From the point of view of "descriptive dynamics" however (a term likely coined by Kechris), we are not just interested in the structure up to measure zero sets, but rather what can be done everywhere in a definable (say Borel) manner. It is also of interest to restrict from Borel to continuous in many questions, that is, asking what can be done in continuous manner leads to interesting questions as well.

In the rest of this section we first give a brief (and selective) background on some results concerning Borel equivalence relations, and then describe some recent work on some problems in this area. We are particularly interested in problems concerning the combinatorial structure of these quotient spaces. We also mention some questions which arise when going past Borel equivalence relations to consider general equivalence relations in determinacy models.

If G is a group acting on the Polish space X, then there is an equivalence relation E_G, the *orbit equivalence relation*, associated to the action: $x \; E_G \; y$ iff $\exists g \in G \; (g \cdot x = y)$. The case of interest is when G is a Polish group (a topological group which is a Polish space in the group topology), and G acts in a Borel way on X (that is, the relation $R(g, x, y) \leftrightarrow g \cdot x = y$ is Borel). An important special of this is when G is a countable discrete group, in which case E_G is a *countable* Borel equivalence relation, that is, all of the E_G classes are countable. In the case of a general Polish group, the relation E_G need only be $\boldsymbol{\Sigma}^1_1$, though it is a fact that all of the individual orbits $[x]_{E_G}$ are Borel. When G is countable, E_G is Borel. Given any countable group G, a natural action is the (left) *shift action* of G on 2^G defined by $g \cdot x(h) = x(g^{-1}h)$. The is a natural action and is also important as it is essentially a universal action of G (we refer the reader to Dougherty, Jackson, and A. S. Kechris [1994] for details).

The theory largely splits in two directions: the case of general (uncountable) Polish groups, and the case where G is countable. Both directions are interesting. For example, the Polish group S_∞ of permutation of ω has a natural action, the *logic action* on the space of countable models of first-order theories (which can viewed as a Polish space). Various important questions in model theory/logic can be phrased as question about this Polish group action. One such question is the well-known Vaught conjecture on the number of models of a first-order theory (that is, is either countable or of size c), which can rephrased as a question about this action (we refer the reader to A. S. Kechris and H. Becker [1996] for details).

For the rest of this section we will focus on the case of countable G, which is illustrative of the general case and includes many cases of interest, particularly in relation to dynamics, ergodic theory, and some aspects of descriptive set theory (we note here that the degree notions of descriptive set theory such as Turing degree, arithmetical degrees, Δ^1_1 degrees, etc., all give countable equivalence relations). We refer the reader to Dougherty, Jackson, and A. S. Kechris [1994], Jackson, A. Kechris, and Louveau [2002], and A. S. Kechris and Miller [2004] for more general background.

The Feldman-Moore theorem Feldman and Moore [1977] is fundamental to the study of countable Borel equivalence relations.

Theorem 10. *Let E be a countable Borel equivalence relation on the Polish space X. then there is countable group G and a Borel action $G \curvearrowright X$ of G on X such that $E = E_G$.*

Thus, we may approach the study of countable Borel equivalence relations "group by group," starting with the algebraically simplest groups and progressing through groups of increasing complexity. Finite groups only generate finite equivalence relations, and these are smooth since there is a Borel linear order on X which we can use to select the least element from each class. The simplest infinite group is $\mathbb{Z}$. Since E_0 is given by a Borel action of $\mathbb{Z}$, these relations need not be smooth. A basic result of Slaman-Steel identifies these as the *hyperfinite* equivalence relations.

Definition 11. A countable Borel equivalence relation E is hyperfinite if $E = \bigcup_n E_n$ is the increasing union of finite equivalence relations (that is, each E_n class is finite).

The Slaman-Steel theorem (see Dougherty, Jackson, and A. S. Kechris [1994]) says that a countable Borel equivalence relation is hyperfinite iff there is a Borel ordering $<_X$ on X such restricted to each class, $<_X \restriction [x]$ is either finite or order-isomorphic to $\mathbb{Z}$. That is, we have in a uniform Borel manner put the structure of a $\mathbb{Z}$ ordering onto each equivalence class.

The fundamental notion in the theory of Borel equivalence relations is the notion of a reduction: we say $(X, E) \leq (Y, F)$ if there is a Borel $f : X \to Y$ such that for all $x, y \in X$, $x \mathrel{E} y \leftrightarrow f(x) \mathrel{F} f(y)$. This is saying that have in a definable way (in this case a Borel way) an injection from the quotient space X/E to Y/F. In other words, this corresponds to saying that X/E has a definable cardinality no larger than that of Y/F. Again, "Borel" can be viewed as a stand-in for other notions of definability; we could consider models of determinacy and allow arbitrary functions f. The Cantor-Schroeder-Bernstein theorem applies here, so if $(X, E) \leq (Y, F)$ and $(Y, F) \leq (X, E)$, then the quotient spaces are in bijection. A result of Dougherty, Jackson, and A. S. Kechris [ibid.] says that all non-smooth hyperfinite equivalence relations are Borel bi-reducible, so they all the same definable cardinality. A central result in the subject is the *Harrington-Kechris-Louveau* dichotomy theorem (see Harrington, A. S. Kechris, and Louveau [1990]). This theorem states that if the Borel equivalence relation (X, E) is not smooth, then $E_0 \leq (X, E)$. That is, there is nothing between the trivial (smooth) relations and the hyperfinite relations. In other words, E_0 is the smallest definable cardinal past those given as subsets of a Polish space (among those representable as Borel equivalence relations).

Two general questions are immediately suggested. The first involves understanding the definable cardinalities of these quotient spaces. That is, determine the structure of the reducibility relation among the family of Borel equivalence relations (or within the countable Borel equivalence relations). The second questions concerns the hyperfinite equivalence relations: which countable groups G generate hyperfinite equivalence relations. That is, which groups G have the property that if $G \curvearrowright X$ is a Borel action of G on the Polish space X, then the orbit equivalence relation E_G is hyperfinite? This hyperfiniteness question was first raised explicitly by Kechris and Weiss. The Connes-Feldman-Weiss theorem answers this question in the ergodic theory/dynamics perspective, that is, up to measure 0 sets with respect to an invariant probability measure μ on X. Their theorem says that if G is *amenable* then, up to a measure 0 set, the action is hyperfinite, and conversely, if all the Borel actions of G are hyperfinite up to a measure 0 set for some such measure, then G is amenable. So, if G is non-amenable then there are Borel actions of G which are not (everywhere) hyperfinite. The other direction is far from clear, and is an important open problem in the area.

Concerning the first problem, a result of Dougherty, Jackson, and A. S. Kechris [1994] shows there is a "largest" countable Borel equivalence relation in the sense that every countable Borel equivalence relations reduces to it. This is given by the shift action of the group F_2 on 2^{F_2} (F_2 here is the free group on 2 generators). While it is not too difficult to show that there are incomparable Borel equivalence relations, the corresponding result for countable Borel equivalence relations was open for a significant time. Finally, A. S. Kechris and Adams [2000] resolved this problem using techniques from Zimmer's superrigidity theory in ergodic theory. They showed that there is a large family (of size continuum) of pairwise incomparable countable Borel equivalence relations. This result was strengthened by Hjorth. In Miller [n.d.] an elegant simplified presentation of some of these results can be found.

We mention the best currently known results on the hyperfiniteness problem. First, Weiss showed (unpublished, but see Jackson, A. Kechris, and Louveau [2002]) the following.

Theorem 12 (Weiss)**.** *All equivalence relations generated by a Borel action of* $\mathbb{Z}^n$ *are hyperfinite.*

Next, Gao and the author extended the result to general abelian groups:

Theorem 13 (Gao and Jackson [2015])**.** *All equivalence relations generated by a Borel action of a countable abelian group are hyperfinite.*

The method used to prove Theorem 13 is quite different from that of Theorem 12. Both proofs employ heavily the use of certain *marker structures* on the equivalence relation. By a marker structure we mean a Borel set $M \subseteq X$ which is complete (meets every equivalence class) and co-complete (its complement is complete). For the proofs, it is necessary to create marker structures with certain delicate geometric properties. Thus, some of the fundamental questions in this area are closely connected with the question of what types of marker structures we can put (in a Borel manner) on the equivalence relation. For the proof of Theorem 13, the notion of an *orthogonal* marker structure was introduced. This roughly says that the marker points M give a decomposition of the points in an equivalence class into rectangular regions such that any two parallel faces of nearby regions are separated by a certain fixed positive fraction of the side lengths. The technology used in this proof has other applications. For example, it allows us to show that there is a *continuous* embedding from $2^{\mathbb{Z}^n}$ (with the shift action) into E_0 (the fact that there is a Borel action follows from the shift action on $2^{\mathbb{Z}^n}$ being hyperfinite). It also allows us to show that the Borel chromatic number of $F(2^{\mathbb{Z}^n})$ is 3 (we discuss this more below), and answer other combinatorial structuring questions.

Theorem 13 was extended further by Schneider and Seward who extended the result to nilpotent groups, and in fact showed the following.

Theorem 14 (Schneider and Seward [n.d.])**.** *All equivalence relations generated by the action of a locally nilpotent group are hyperfinite.*

By a an important result of Gromov in geometric group theory, the class of finitely generated groups which have a nilpotent subgroup of finite index (the virtually nilpotent groups) coincides with the class of finitely generated groups of polynomial growth. We note that Theorem 14 for the case of finitely generated nilpotent groups (or finitely generated virtually nilpotent groups) was known previously, a result of Jackson, A. Kechris, and Louveau [2002]. This suggested the possibility that polynomial growth was the barrier to extending these hyperfiniteness results. However, in recent as yet unpublished work, Conley, Marks, Seward, Tucker-Drob, and the author have shown that there are finitely generated solvable, non-nilpotent (so not of polynomial growth) groups all of whose Borel actions are hyperfinite. Whether these arguments can be made to extend to all elementary amenable groups, or even all amenable groups, is not yet known.

Aside from the above questions concerning the cardinalities of the quotient spaces X/E, we are interested in questions about the combinatorial structure of these sets. We can ask these types of question at either the definable level, where we usually use "Borel" as a representative case, or at the topological level. Roughly speaking, in the latter case, we require the types of structures we are considering to be given in a continuous manner. As we said above, the hyperfiniteness arguments require certain types of marker structures, but there are many other kinds of structuring questions we can ask.

The notion of a continuous or Borel "structuring" of the countable Borel equivalence relation E can be made precise in a natural manner. If $\mathfrak{L} = (c_i, R_i, f_i)$ is a language of first-order logic, by an $\mathfrak{L}$-structuring of E we mean an assignment $[x] \mapsto \mathfrak{A}_x$ of $\mathfrak{L}$-structures $\mathfrak{A}_x$ to the equivalence classes of E, where the domain of the structure $\mathfrak{A}_x$ is the equivalence class $[x]$. If $E = E_G$ for some action of the group G, then we frequently also assume that there are unary function symbols f_g in the language $\mathfrak{L}$ for each group element $g \in G$ (these are intended to represent the function $f_g(x) = g \cdot x$). The notion of the structuring being Borel (or continuous) is defined in a natural manner (e.g., for each n-ary relation symbol R_i of $\mathfrak{L}$, the relation

$$R(x_1, \dots, x_n) \leftrightarrow x_1 \; E \; x_2 \cdots E \; x_n \wedge \mathfrak{A}_{[x_1]}(x_1, \dots, x_n)$$

is a Borel (or clopen) relation on X^n. We can then, for example, ask if Borel or continuous structurings of E exists with the structures $\mathfrak{A}_x$ satisfying certain properties (for example, if they satisfy a certain formula of first-order, or higher-order, logic).

Many types of interesting combinatorial questions can be phrased as instances of structuring questions. Consider a fixed countable group G. Given actions $G \curvearrowright X$ and $G \curvearrowright Y$ generating equivalence relations E_X and E_Y, we say $f : X \to Y$ is *equivariant* if f commutes with the actions, that is, $f(g \cdot x) = g \cdot f(x)$ for all $x \in X$. A one-to-one equivariant map is necessarily a reduction from E_X to E_Y. By a subshift of finite type we mean a closed, invariant (under the shift action) $Y \subseteq k^G$ for some $k \in \mathbb{N}$ which is defined by a finite set $p_1, \dots, p_t$ of "forbidden" patterns. Here a pattern is partial function $p : G \to k$ with finite domain. Then $y \in k^G$ is in the subshift Y determined by the p_i (with say $D_i = \mathrm{dom}(p_i)$) if for all $g \in G$, the function $p_g : D_i \to k$ given by $p_g(h) = g \cdot y(h)$ is not equal to any of the p_i. Asking if there is an equivariant map from E_X to the subshift Y is an instance of a structuring question. Subshift questions of this form are themselves quite general and include several interesting types of questions. We consider a few of these types of questions and some recent results concerning them.

If G is a marked group, that is, comes with a distinguished set of generators S (which does include the identity e), then there is a graphing $\Gamma(E_G)$ of the orbit equivalence relation E_G for any action $G \curvearrowright X$ given by $x \Gamma(E_G) y$ iff $\exists s \in S \; (s \cdot x = y \vee s \cdot y = x)$. If the action is free, then on each equivalence class this graphing is isomorphic to the

Cayley graph associated to (G, S). The *Borel chromatic number*, $\chi_b(E_G)$ of the equivalence relation is the least cardinal k such that there is a Borel map $c\colon X \to k$ which is a chromatic coloring of the graph $\Gamma(E_G)$. We likewise define the *continuous chromatic number* $\chi_c(E_G)$, using continuous functions f. The study of definable chromatic numbers was initiated by Kechris, Solecki, and Todorcevic in A. S. Kechris, Solecki, and Todorcevic [1999]. One of their basic results is that the Borel chromatic number satisfies $\chi_b(E_G) \leq d+1$, where d is the vertex degree of the Cayley graph, for any E_G generated by a free action of G. We refer to the determination of $\chi_c(E_G)$ and $\chi_b(E_G)$ as the chromatic number problem. This is an instance of the more general *subshift problem*, which is determine for which subshifts $Y \subseteq k^G$ (determined by k and the patterns $p_1, \dots, p_t$) there is a continuous or Borel equivariant map from E_G to Y. Another instance of the subshift problem is the *graph homomorphism problem*. Given a countable graph Γ, this problem is to determine whether there is a continuous or Borel graph homomorphism from $\Gamma(E_G)$ to Γ. Finally, we mention the *tiling problem*. By a tile we mean a finite set $T \subseteq G$. Given a finite set $T_1, \dots, T_\ell$ of tiles, the tiling problem asks whether there is a continuous or Borel tiling of E_G. By this we mean Borel sets $A_i \subseteq X$ such that the sets $\{T_i \cdot g\colon g \in A_i\}$ partition X (a "continuous" tiling means that the A_i are clopen sets in X). There are many other types of structuring questions one can ask, but these serve as test questions for the type of definable structures we can put on the equivalence classes. While these questions are of interest for general countable groups, let us now restrict our attention to simpler groups.

Consider the groups $G = \mathbb{Z}^n$. As we said above, all of these groups induce only hyperfinite actions. Nevertheless, structuring questions about the equivalence relations generated by actions of these groups are non-trivial. Perhaps even more surprising, given the fact that all of these shift spaces $2^{\mathbb{Z}^n}$ continuously embed into E_0, is that some continuous structuring questions have answers that depend on n.

We mentioned above the method of orthogonal markers, which has been used in recent hyperfiniteness proofs. This method is normally used in a "positive" sense, that is, to produce Borel structurings on various types in the equivalence relations E_G. Another method which has been used to obtain negative results in the continuous setting is the method of *hyperaperiodic* points. The notion of hyperaperiodic point was introduced by Gao, Seward, and the author in Gao, Seward, and Jackson [2009] and also independently by Glasner and Uspenskij. Consider the shift space 2^G. We say $x \in 2^G$ is a hyperaperiodic point if $\overline{[x]} \subseteq F(2^G)$, that is, the closure of the orbit of x lies entirely in the free part of 2^G. This definition can be reformulated as a purely combinatorial property of x. Namely, $x \in 2^G$ is hyperaperiodic iff it satisfies the following: for any $s \in G$ with $s \neq e$, there is a finite $T \subseteq G$ such that

$$\forall g \in G\ \exists t \in T\ (x(gt) \neq x(gst)).$$

This combinatorial property is sometimes referred to as x being a "2-coloring." Hyperaperiodic elements are easy to construct for simple groups such as $\mathbb{Z}^n$, however the following result of Gao, Seward, and Jackson [ibid.] states that they exist for any countable group.

Theorem 15 (Gao, Seward, and Jackson [ibid.]). *For every countable group G there is an $x \in 2^G$ which is a hyperaperiodic point.*

Hyperaperiodic points are useful since $\overline{[x]}$ is a compact set contained within the free part of 2^G, and this permits certain compactness arguments. However, to prove some more delicate results it is necessary to construct hyperiodic points with various addition properties. To illustrate the use of the orthogonal marker and hyperaperiodic element arguments, consider the Borel and continuous chromatic number problems for $F(2^{\mathbb{Z}^n})$. An easy category argument shows that $\chi_b(F(2^{\mathbb{Z}^n}) > 2$ for all n. Also, any easy argument using the existence of clopen marker sets (see Gao and Jackson [2015]) which are roughly d-spaced (for any $d > 1$) shows that $\chi_c(F(2^{\mathbb{Z}})) \leq 3$. It follows that in the $n = 1$ case we have $\chi_b(F(2^{\mathbb{Z}})) = \chi_c(F(2^{\mathbb{Z}})) = 3$. For $n \geq 2$, the arguments require the new methods. The result, from Gao, Jackson, Krohne, and Seward [n.d.], is the following.

Theorem 16 (Gao, J, Krohne,Seward). *For any $n \geq 2$ we have:*

$$3 = \chi_b(F(2^{\mathbb{Z}^n})) < \chi_c(F(2^{\mathbb{Z}^n})) = 4.$$

This result has two points of interest. First, it shows a difference between the dimension $n = 1$ and $n \geq 2$ cases, even though both equivalence relations are hyperfinite. Second, it shows a difference between the continuous and Borel versions of the question.

The proof of Theorem 16 was first first accomplished by the construction of a particular hyperaperiodic point. The basic idea was to construct $x \in 2^{\mathbb{Z}^n}$ with certain periodicity requirements in one direction, but yet keeping the point hyperaperiodic. This is possible as $n \geq 2$. Later, Gao, Krohne, Seward, and the author proved a general theorem applicable to general subshift questions. The theorem (see Gao, Jackson, Krohne, and Seward [ibid.]) reduces the subshift question for $F(2^{\mathbb{Z}^n})$ down to a question about a family of finite graphs. Consider the case $n = 2$. For each $1 \leq n < p, q$ we define a finite graph $\Gamma_{n,p,q}$. The graph is obtained by starting with 12 individual "grid-graphs." by a grid-graph we mean a graph which is isomorphic to a finite rectangular region of $\mathbb{Z}^2$ with edges inherited from the Cayley graphing of $\mathbb{Z}^2$. Certain vertices are identified among the vertices in these grid-graph, and the quotient graph is $\Gamma_{n,p,q}$. To give the reader a feel for the construction we show three of the grid graphs in Figure 4 (each of the other graphs comprising $\Gamma_{n,p,q}$ is similar to one of these). In each of these grid-graphs, certain rectangular subregions are marked with labels such as $R_\times$, R_a, etc. In the graph $\Gamma_{n,p,q}$ the corresponding points within regions with the same label are identified. For example, the upper-left points of each $R_\times$ region are identified in forming $\Gamma_{n,p,q}$.

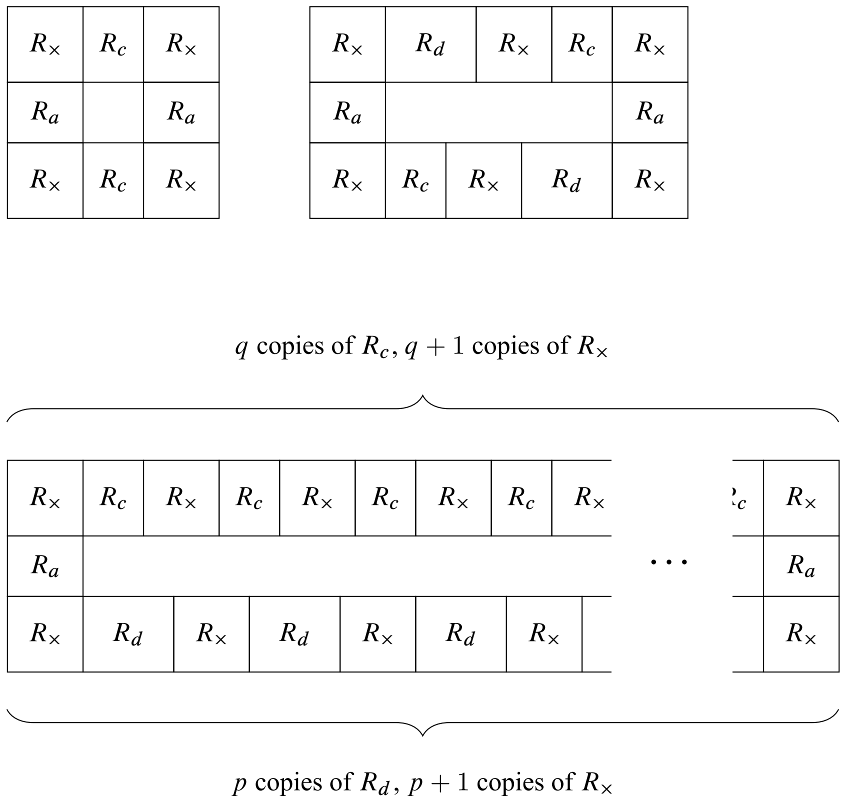


Figure 4: The grid-graphs in $\Gamma_{n,p,q}$.

The following result of Gao, Jackson, Krohne, and Seward [n.d.] shows that a subshift question for $F(2^{\mathbb{Z}^n})$ reduces to question about the graph $\Gamma_{n,p,q}$.

Theorem 17. *Let $Y \subseteq k^{\mathbb{Z}^2}$ be a subshift of finite type described by $(k; p_1, \dots, p_t)$. Then the following are equivalent.*

1. *There is a continuous, equivariant map $f\colon F(2^{\mathbb{Z}^2}) \to Y$.*
2. *There are positive integers n, p, q with $n < p, q$, $(p, q) = 1$, and $n \geq \max\{a_i, b_i : dom(p_i) = [0, a_i) \times [0, b_i)\} - 1$ and a $g\colon \Gamma_{n,p,q} \to k$ which respects Y.*
3. *For all $n \geq \max\{a_i, b_i : dom(p_i) = [0, a_i) \times [0, b_i)\} - 1$, for all sufficiently large p, q with $(p, q) = 1$ there is a $g\colon \Gamma_{n,p,q} \to k$ which respects Y.*

In this theorem, when we say $g\colon \Gamma_{n,p,q} \to k$ respects the subshift y we mean that in any $a_i \times b_i$ rectangular subregions R of one of the grid-graphs forming $\Gamma_{n,p,q}$, $g \restriction R$ is not equal to p_i. In other words, we can find a continuous equivariant map from $F(2^{\mathbb{Z}^2})$ into the subshift $Y \subseteq k^{\mathbb{Z}^2}$ iff we can find such a map from $\Gamma_{n,p,q} \to k$ for some p, q with $(p, q) = 1$ (equivalently, if we can such maps g for all sufficiently large p, q with $(p, q) = 1$).

Using this result, a number of subshift questions can be answered for $F(2^{\mathbb{Z}^n})$. Moreover, some general results about the decidability of the subshift problem in general can be shown which highlight a key difference between the dimension $n = 1$ and $n \geq 2$ cases. A subshift Y is coded by a finite sequence $(k; p_1, \ldots, p_t)$, which can be viewed as an integer. Let Y_m be the subshift coded by the integer m. Consider the set $S(n)$ of $m \in \omega$ such that there is a continuous, equivariant map from $F(2^{\mathbb{Z}^n})$ to Y_m. From Theorem 17 it follows that for each n the set $S(n)$ is computably enumerable, that is, is a Σ^0_1 set. The question we consider is whether this set is actually computable (i.e., a Δ^0_1 set). We have the following result of Gao, Jackson, Krohne, and Seward [ibid.].

Theorem 18. *For $n = 1$, the subshift problem is decidable, that is, $S(1)$ is computable. For $n \geq 2$ the subshift problem is not computable.*

Theorem 18 shows a remarkable difference between the shift actions of $\mathbb{Z}$ and $\mathbb{Z}^n$ for $n \geq 2$. In Gao, Jackson, Krohne, and Seward [ibid.] it is further shown that even the specific graph homomorphism problem for $F(2^{\mathbb{Z}^n})$ is not computable for $n \geq 2$.

The above results are for the shift actions of the groups $\mathbb{Z}^n$. Let us mention a result of a similar flavor but for a completely different class of groups. this result, obtained by Marks [2016] concerns the free product of groups. The result is:

Theorem 19 (Marks). *If G, H are finitely generated marked groups, then*

$$\chi_b(F(2^{G*H})) \geq \chi_b(F(2^G) + \chi_b(F(2^H)) - 1.$$

where $F(2^G)$ denotes the free part of the shift action of G on 2^G and $G * H$ denotes the free product of the groups G and H (the statement of Theorem 19 above actually incorporates an improvement of the result due to Seward and Tucker-Drob). What is interesting is that Marks' method in proving this result involves games and Borel determinacy (a result due to Martin). This surprising result introduces yet another new technique into the subject.

In this section we have been mainly concerned with objects given by Borel equivalence relations on Polish spaces. Although Borel is frequently taken as a representative of definable, let us finally return to considering general sets in a model determinacy. Woodin [2006] has shown an interesting result about cardinalities in determinacy models which shows that the exact determinacy hypotheses assumed may be important. Woodin shows that assuming $\mathsf{AD}_{\mathbb{R}}$, the axiom of real game determinacy (which is considerably stronger than AD or AD^+), there are exactly 5 cardinals below the set ω_1^ω. He also shows that there are more than 5 cardinals below this set if one assumes $\mathsf{AD} + \neg\mathsf{AD}_\mathbb{R}$. This surprising result shows that for some sets, questions about their definable structure may depend on the background axioms assumed.

The results we have discussed show that a very rich theory of definable sets is emerging, and is connected with many other areas of mathematics. We believe this will continue to be an interesting and fruitful line of investigation.

References

Arthur Apter, Steve Jackson, and Benedikt Löwe (2013). "Cofinality and measurability of the first three uncountable cardinals". *Transactions of the American Mathematical Society* 365, pp. 59–98 (cit. on p. 50).

Howard S. Becker and S. Jackson (2001). "Supercompactness within the projective hierarchy". *Journal of Symbolic Logic* 66.2, pp. 658–672 (cit. on p. 51).

Randall Dougherty, Steve Jackson, and Alexander S. Kechris (1994). "The Structure of Hyperfinite Borel Equivalence Relations". *The Transactions of the American Mathematical Society* 341, pp. 193–225 (cit. on pp. 53–55).

Jacob Feldman and Calvin Moore (1977). "Ergodic equivalence relations, cohomology, and von Neumann algebras, I". *Transactions of the American Mathematical Society* 234.2, pp. 289–324 (cit. on p. 54).

Su Gao and Steve Jackson (2015). "Countable Abelian group actions and hyperfinite equivalence relations". *Inventiones mathematicae* 201.1, pp. 309–383 (cit. on pp. 56, 59).

Su Gao, Steve Jackson, Ed Krohne, and Brandon Seward (n.d.). "Continuous combinatorics of Abelian group actions" (cit. on pp. 59–61).

Su Gao, Brandon Seward, and Steve Jackson (2009). "A coloring property for countable groups". *Mathematical proceedings of the Cambridge Philosophical Society* 147.3, pp. 579–592 (cit. on pp. 58, 59).

L. Harrington, A. S. Kechris, and A. Louveau (1990). "A Glimm-Effros dichotomy for Borel equivalence relations". *Journal of the American Mathematical Society* 3.4, pp. 903–928 (cit. on p. 55).

Steve Jackson (1988). "AD and the projective ordinals". In: *Cabal Seminar 81–85*. Vol. 1333. Lecture Notes in Math. Berlin: Springer, pp. 117–220 (cit. on pp. 47, 49).

– (1989). "AD and the projective ordinals". *Bulletin of the American Mathematical Society* 21, pp. 77–81 (cit. on p. 48).

– (1999). "A computation of $\boldsymbol{\delta}^1_5$". *Memoirs of the A.M.S.* 140.670, pp. 1–94 (cit. on pp. 47, 49).

– (2001). "The weak square property". *Journal Symbolic Logic* 66.2, pp. 640–657 (cit. on p. 51).
– (2010). "Structural consequences of AD". In: *Handbook of set theory*. Vol. 3. Springer, pp. 1753–1876 (cit. on pp. 47–49).
– (2011). "Regular cardinals without the weak partition property". In: *The Cabal seminar volume II, Lecture notes in Logic*. Vol. 35. Association for Symbolic Logic, pp. 509–518 (cit. on p. 50).
Steve Jackson, Alexander Kechris, and Alain Louveau (2002). "Countable Borel equivalence relations". *Journal of Mathematical Logic* 2, pp. 1–80 (cit. on pp. 54–56).
Steve Jackson, Richard Ketchersid, Farmer Schlutzenberg, and W. Hugh Woodin (2014). "AD and Jónsson cardinals in $L(\mathbb{R})$". *The Journal of Symbolic Logic* 79.4, pp. 1184–1198 (cit. on p. 51).
Steve Jackson and Farid Khafizov (2016). "Descriptions and cardinals below $\boldsymbol{\delta}^1_5$". *The Journal of Symbolic Logic* 84.4, pp. 1177–1224 (cit. on p. 49).
Steve Jackson and Löwe (2013). "Canonical measure assignments". *The Journal of Symbolic Logic* 78.2, pp. 403–424 (cit. on p. 49).
A. S. Kechris and S. Adams (2000). "Linear algebraic groups and countable Borel equivalence relations". *Journal of the American Mathematical Society* 13.4, pp. 909–943 (cit. on p. 55).
A. S. Kechris and B. D. Miller (2004). *Topics in orbit equivalence*. Springer (cit. on p. 54).
A. S. Kechris, S. Solecki, and S. Todorcevic (1999). "Borel chromatic numbers". *Advances in Mathematics* 141 (1), pp. 1–44 (cit. on p. 58).
Alexander S. Kechris (1978). "AD and the projective ordinals". In: *Cabal seminar 76–77*. Vol. 689. Lecture Notes in Math. Berlin: Springer, pp. 91–132 (cit. on pp. 48, 49).
Alexander S. Kechris and Howard Becker (1996). *The descriptive set theory of Polish group actions*. Vol. 232. London Mathematical Society Lecture Notes Series. Cambridge University Press, pp. 1–152 (cit. on p. 54).
Andrew Marks (2016). "A determinacy approach to Borel combinatorics". *Journal of the American Mathematical Society* 29, pp. 579–600 (cit. on p. 61).
B. D. Miller (n.d.). "Incomparable treeable equivalence relations". *Journal of Mathematical Logic* 12.1:1250004 () (cit. on p. 55).
Yiannis N. Moschovakis (1980). *Descriptive set theory*. Vol. 100. Studies in logic. North–Holland (cit. on pp. 46–48).
S. Schneider and B. Seward (n.d.). "Locally nilpotent groups and hyperfinite equivalence relations". arXiv: `1308.5853` (cit. on p. 56).
John R. Steel (1995). "$\mathrm{HOD}^{L(\mathbb{R})}$ is a core model below Θ". *Bull. Symbolic Logic* 1.1, pp. 75–84 (cit. on p. 51).
W. Hugh Woodin (2006). "The cardinals below $|[\omega_1]^{<\omega_1}|$". *Annals of pure and applied logic* 140, pp. 161–232 (cit. on p. 61).

Received 2018-03-02.

Stephen Jackson
Department of Mathematics
University of North Texas
Denton, TX, 76203
Stephen.Jackson@unt.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (63–78)

DECIDABILITY IN LOCAL AND GLOBAL FIELDS

Jochen Koenigsmann

Abstract

This lecture highlights some recent advances on classical decidability issues in local and global fields.

Contents

1 Introduction

In 1900, at the International Congress of Mathematicians in Paris, Hilbert presented his celebrated and influential list of 23 mathematical problems (Hilbert [1900]). One of them is

Hilbert's 10th Problem (H10) Find an algorithm which gives on INPUT any $f(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$

$$\text{OUTPUT} \begin{cases} \textbf{YES} & \text{if } \exists \overline{x} \in \mathbb{Z}^n \text{ such that } f(\overline{x}) = 0 \\ \textbf{NO} & \text{else} \end{cases}$$

Hilbert did not ask to prove that there is such an algorithm. He was convinced that there should be one, and that it was all a question of producing it — one of those instances of

MSC2010: primary 11U05; secondary 03B25, 11F85, 12J10, 12L05.

Hilbert's optimism reflected in his famous slogan '*wir müssen wissen, wir werden wissen*' ('*we must know, we will know*'). As it happens, Hilbert was too optimistic: after previous work since the 50's by Martin Davis, Hilary Putnam and Julia Robinson (cf., e.g., Davis, Putnam, and J. Robinson [1961]), Yuri Matiyasevich showed in 1970 that there is no such algorithm (Y. V. Matiyasevich [1970b]). The key result here is the most remarkable so-called DPRM-Theorem that every (algorithmically) listable set of integers is *diophantine*, i.e., first-order definable in the language of rings $L_{ring} := \{+, \times, 0, 1\}$ by an *existential* formula.

The original formulation of Hilbert's 10th Problem was weaker than the standard version above in that he rather asked *'Given a polynomial f, find an algorithm ...'*. So maybe you could have different algorithms depending on the number of variables and the degree. Yet it is even possible to find a single polynomial for which no such algorithm exists — this is essentially because there are universal Turing Machines.

One should, however, mention that, in the special case of $n = 1$, that is, for polynomials in one variable, there is an easy algorithm: if, for some $x \in \mathbb{Z}$, $f(x) = 0$ then $x \mid f(0)$; hence one only has to check the finitely many divisors of $f(0)$. Similarly, by the effective version of the Hasse-Minkowski-Local-Global-Principle for quadratic forms and some extra integrality considerations, one also has an algorithm for polynomials in an arbitrary number of variables, but of total degree ≤ 2. And, even if there is no general algorithm, it is one of the major projects of computational arithmetic geometry to exhibit other families of polynomials for which such algorithms exist.

Let us point in a different direction of generalizing Hilbert's 10th Problem, namely, generalizing it to rings other than $\mathbb{Z}$: If R is an integral domain, there are two natural ways of generalizing **H10**:

H10/R = **H10** with the 2nd occurrence of $\mathbb{Z}$ replaced by R
H10$^+$/R = **H10** with both occurrences of $\mathbb{Z}$ replaced by R

Observation 1.1. *Let R be an integral domain whose field of fractions does not contain the algebraic closure of the prime field ($\mathbb{F}_p$ resp. $\mathbb{Q}$). Then*

$$\begin{aligned} \mathbf{H10}/R \text{ is solvable } &\Leftrightarrow \mathtt{Th}_{\exists^+}(R) \text{ is decidable} \\ \mathbf{H10}^+/R \text{ is solvable } &\Leftrightarrow \mathtt{Th}_{\exists^+}(\langle R; r \mid r \in R\rangle) \text{ is decidable,} \end{aligned}$$

where $\mathtt{Th}_{\exists^+}$ denotes the positive existential theory consisting of existential sentences where the quantifier-free part is a conjunction of disjunctions of polynomial equations (no inequalities).

Note that the language on the right hand side of the 2nd line contains a constant symbol for each $r \in R$.

Proof: '$\Leftarrow$' is obvious in both cases. For '$\Rightarrow$' one has to see that a disjunction of two polynomial equations is equivalent to (another) single equation, and, likewise, for conjunctions: By our assumption we can find some monic $g \in \mathbb{Z}[X]$ of degree > 1 which is irreducible over R. Then, for any polynomials f_1, f_2 over $\mathbb{Z}$ resp. R and for any tuple $\overline{x}$ over R,

$$f_1(\overline{x}) = 0 \vee f_2(\overline{x}) = 0 \iff f_1(\overline{x}) \cdot f_2(\overline{x}) = 0$$
$$f_1(\overline{x}) = 0 \wedge f_2(\overline{x}) = 0 \iff g\left(\frac{f_1(\overline{x})}{f_2(\overline{x})}\right) \cdot f_2(\overline{x})^{\deg g} = 0$$

□

Since in fields, inequalities can be expressed by a positive existential formula ($f(\overline{x}) \neq 0 \leftrightarrow \exists y \;\, f(\overline{x}) \cdot y = 1$), we immediately obtain the following:

Corollary 1.2. *Let K be a field not containing the algebraic closure of the prime field. Then*

$$\textbf{H10}/K \textit{ is solvable } \Leftrightarrow \mathtt{Th}_{\exists}(K) \textit{ is decidable.}$$

In fact, the same is true for $\mathcal{O}_K$, the ring of integers of a number field K:

Observation 1.3. *If K is a number field,*

$$\mathcal{O}_K \models \forall x[x \neq 0 \leftrightarrow \exists y \; x \mid (2y-1)(3y-1)].$$

Hence $\mathtt{Th}_{\exists}(\mathcal{O}_K) = \mathtt{Th}_{\exists^+}(\mathcal{O}_K)$.

One of the biggest open questions in the area is

Question 1.4. *Is* **H10**$/\mathbb{Q}$ *solvable?*

Let us recall that, by the ground breaking work of Kurt Gödel, the full first order theory of $\mathbb{Z}$ is undecidable, so there is no algorithm which decides, on INPUT any first-order L_{ring}-sentence ϕ, whether or not ϕ holds in $\mathbb{Z}$ (cf. Gödel [1931]). J. Robinson [1949] managed to find an L_{ring}-first-order definition of $\mathbb{Z}$ in $\mathbb{Q}$, thus showing, via Gödel's Theorem, that the full first-order theory of $\mathbb{Q}$ is also undecidable. If one had an *existential* first-order L_{ring}-formula defining $\mathbb{Z}$ in $\mathbb{Q}$ then one could, via Matiyasevich's Theorem, conclude that Hilbert's 10th Problem over $\mathbb{Q}$ is also unsolvable. However, the best we have at the moment (in terms of logical complexity) is a *universal* formula for $\mathbb{Z}$ in $\mathbb{Q}$ (cf. Theorem 3.1 below).

Hilbert's 10th Problem for the ring of integers of a number field (that is, a finite extension of $\mathbb{Q}$ — they are the *global* fields of characteristic 0) has been shown to be unsolvable in several cases, the general case could sofar only be proven modulo a (widely believed) conjecture regarding elliptic curves (see section 3.2).

For global fields of positive characterisitc, that is, for finite extensions of the rational function field $\mathbb{F}_p(t)$ over the finite field $\mathbb{F}_p$ in one variable t, Hilbert's 10th Problem, again, has no solution (cf. section 3.3).

Many of the results obtained for global fields rely heavily on results and techniques developed for *local fields*. Local fields are defined to be fields F which are locally compact with respect to the topology induced by some non-trivial absolute value on F. It turns out that local fields are precisely the completions of global fields (w.r.t such absolute values) and they are classified as follows: the *archimedean* local fields are just the field $\mathbb{R}$ of real numbers and the field $\mathbb{C}$ of complex numbers; the non-archimedean local fields of characteristic 0 are precisely all finite extensions of $\mathbb{Q}_p$, the field of *p-adic numbers*, where p is any rational prime; and the non-archimedean local fields of positive characteristic p are precisely the finite extensions of $\mathbb{F}_p((t))$, the field of formal Laurent series over the field $\mathbb{F}_p$ with p elements. For the non-archimedean local fields, the absolute value is induced by a canonical valuation, which is the p-adic valuation on $\mathbb{Q}_p$ and the t-adic valuation on $\mathbb{F}_p((t))$, and these valuations extend uniquely to all finite extensions, a property of valuations called *henselian*.

All decidability issues for the two archimedean local fields have been settled by Tarski in the 1930s: The full first order theory of $\mathbb{R}$ and of $\mathbb{C}$ is decidable (and hence, in particular, Hilbert's 10th Problem is solvable for those two fields).

The decidability of $\mathbb{Q}_p$ was proved independently by Ax and Kochen [1965] and by Eršov [1965]. They effectively axiomatized $\mathbb{Q}_p$ as a henselian valued field of characteristic 0 whose residue field is $\mathbb{F}_p$, whose value group is a $\mathbb{Z}$-group (so elementarily equivalent to the ordered abelian group of integers) such that the value of p is minimal positive. And there are similar axiomatizations for all finite extensions of $\mathbb{Q}_p$ (see Prestel and Roquette [1984] for a general treatment of p-adic fields).

Since those results of Ax-Kochen and Ershov in 1965 it has been a big open problem whether the theory of $\mathbb{F}_p((t))$ is decidable as well. Recently major progress has been made on this problem which we will discuss in section 2 below.

There are several important infinite extension of local and global fields for which decidability issues are of great interest, too. For example, the field $\mathbb{Q}^{ab}$, the maximal Galois extension of $\mathbb{Q}$ with an abelian Galois group which, by the famous Kronecker-Weber Theorem, is just the field obtained from $\mathbb{Q}$ by adjoining all roots of unity, is not known to be decidable or undecidable. Similarly, one does not know this about $\mathbb{Q}^{solv}$, the maximal Galois extension of $\mathbb{Q}$ with prosolvable Galois group which is obtained from $\mathbb{Q}$ by iteratedly adjoining radicals (n-th roots of elements for arbitrary n). It is an open problem in Field Arithmetic whether or not $\mathbb{Q}^{solv}$ is pseudo-algebraically closed in the sense that every absolutely irreducible curve defined over $\mathbb{Q}^{solv}$ has a $\mathbb{Q}^{solv}$-rational point (Problem 11.5.9(a) in Fried and Jarden [2008]). If this Problem has a positive answer and if the famous Shafarevich Conjecture that the absolute Galois group of $\mathbb{Q}^{ab}$ is a free profinite group is true then one can show that $\mathbb{Q}^{solv}$ is decidable.

In section 4 we will briefly consider two infinite extensions of $\mathbb{Q}_p$ for which there has been recent progress, namely $\mathbb{Q}_p^{ur}$, the maximal unramified extension of $\mathbb{Q}_p$ which turns out to be decidable and model theoretically well behaved, and $\mathbb{Q}_p^{ab}$, the maximal abelian extension of $\mathbb{Q}_p$, for which a promising new suggestion for a first-order axiomatization will be presented.

Some notation from valuation theory: The reader is expected to be acquainted with the basics of valuation theory (cf., e.g., Engler and Prestel [2005]). For a valued field (K, v), the valuation ring will be denoted by $\mathcal{O}_v$, the residue field by Kv and the value group by vK.

2 Local fields of positive characteristic

Regarding the question of decidability of the field $\mathbb{F}_q((t))$ of formal Laurent series over a finite field $\mathbb{F}_q$ there have been two recent breakthroughs: one is the result of Anscombe-Fehm that the *existential* L_{ring}-theory of $\mathbb{F}_q((t))$ is decidable (Theorem 2.1). The other is a new promising suggestion for an effective first order axiomatization for $\mathbb{F}_q((t))$ using the notion of *extremal valued fields*.

Throughout this section we will fix q, a power of the rational prime $p > 0$.

2.1 The existential theory of $\mathbb{F}_q((t))$. In Denef and Schoutens [2003], Jan Denef and Hans Shoutens managed to prove that the *existential* theory of $\mathbb{F}_q((t))$ in $L_{ring} \cup \{t\}$, the language of rings augmented by a constant symbol for t, is decidable if one assumes resolution of singularities in positive characteristic. Sylvy Anscombe and Arno Fehm then found a surprisingly elementary unconditional proof for the decidability of the existential L_{ring}-theory of $\mathbb{F}_q((t))$ (see S. Anscombe and Fehm [2016]). More generally they prove the following

Theorem 2.1. *Let (K, v) be an equicharacteristic henselian valued field (so char $K =$ char Kv). Then the existential L_{val}-theory of K is decidable if and only if the existential L_{ring}-theory of the residue field Kv is decidable.*

Here $L_{val} = L_{ring} \cup \{\mathcal{O}\}$ is the language of valued fields, that is, the language of rings augmented by a unary predicate symbol $\mathcal{O}$ for the valuation ring. There are many alternative possibilities for a first order language for valued fields (you could, for example, have a three-sorted language distinguishing the field sort, the residue field sort and the value group sort with additional function symbols for the valuation map and the canonical restriction map to the residue field). But it turns out that all these languages are mutually translatable into each other, so they all have the same expressive power.

Let us point out that, for the question of the decidability of the existential theory of $\mathbb{F}_q((t))$, it makes no difference whether you ask this question about the existential theory in L_{ring} or in L_{val}, because, by the main theorem of W. Anscombe and Koenigsmann [2014], the valuation ring $\mathbb{F}_q[[t]]$ of $\mathbb{F}_q((t))$ is existentially first-order definable in the language of rings. This leads immediately to the following

Corollary 2.2. *The existential L_{ring}-theory of $\mathbb{F}_q[[t]]$ is dedicable.*

So, in other words, Hilbert's 10th Problem has a positive solution both for $\mathbb{F}_q((t))$ and for $\mathbb{F}_q[[t]]$.

A more general result on almost existential definability of henselian valuation rings in valued fields with finite or pseudo-finite residue fields can be found in Cluckers, Derakhshan, Leenknegt, and Macintyre [2013].

Whether or not the existential $L_{ring} \cup \{t\}$-theory of $\mathbb{F}_q((t))$ is decidable (without assuming resolution of singularities) is still open.

2.2 Axiomatizing $\mathbb{F}_q((t))$. The biggest open question in the model theory of valued fields, however, is the question whether the full first-order theory of $\mathbb{F}_q((t))$ is decidable. There have been a number of suggestions of how to axiomatize this field. The most promising suggestion builds on the notion of *extremal* valued fields, originally introduced (though with a 'wrong' definition) by Ershov [2004], then, following a suggestion of Sergei Starchenko, the definition was amended and the 'correct' definition was put forward in Ershov [2009] and in Azgin, Kuhlmann, and Pop [2012]. The suggested axiomatization for $\mathbb{F}_q((t))$ given below first appeared in S. Anscombe and Kuhlmann [2016].

Definition 2.3. *A valued field (K, v) is* extremal *if, for every polynomial $f(X_1, \ldots, X_n) \in K[X_1, \ldots, X_n]$, the set*

$$\{v(f(a_1, \ldots, a_n)) \mid a_1, \ldots, a_n \in \mathcal{O}_v\} \subseteq vK \cup \{\infty\}$$

has a maximal element.

It turns out that extremal valued fields are *algebraically maximal*, that is, for each finite extension $(L, w)/(K, v)$, the fundamental equality '$n = e \cdot f$' holds, where $n = [L : K]$, $e = [wL : vK]$ and $f = [Lw : Kv]$, and so, in particular, extremal fields are henselian. Moreover, their value group is either divisible or a $\mathbb{Z}$-group (elementarily equivalent to the ordered abelian group $\mathbb{Z}$ of integers) and that in the first case the residue field has to be large in the sense of having infinitely many rational points for each algebraic curve with at least one rational point (cf. Pop [1996]).

The axiomatization for $\mathbb{F}_q((t))$ using this notion of extremal valued fields is now very simple:

(1) (K, v) is an extremal valued field of characteristic p,

(2) the value group vK is a $\mathbb{Z}$-group,

(3) the residue field Kv is the field $\mathbb{F}_q$.

It has long been known that the ‘naive’ axiomatization for $\mathbb{F}_q((t))$, where axiom (1) is replaced by just asking (K, v) to be henselian, is not complete.

3 Global fields

3.1 A universal definition for $\mathbb{Z}$ in $\mathbb{Q}$. Hilbert’s 10th problem over $\mathbb{Q}$, i.e., the question whether the existential L_{ring}-theory of $\mathbb{Q}$ is decidable, is still open.

If one had an *existential* (= *diophantine*) definition of $\mathbb{Z}$ in $\mathbb{Q}$ (i.e., a definition by an existential 1st-order $\mathfrak{L}_{ring}$-formula) then the existential theory of $\mathbb{Z}$ would be interpretable in that of $\mathbb{Q}$, and the answer would, by (for short) Matiyasevich’s Theorem, again be no. But it is still open whether $\mathbb{Z}$ is existentially definable in $\mathbb{Q}$.

The earliest 1st-order definition for $\mathbb{Z}$ in $\mathbb{Q}$, due to J. Robinson [1949], can be expressed by an $\forall\exists\forall$-formula of the shape

$$\phi(t):\ \forall x_1 \forall x_2 \exists y_1 \dots \exists y_7 \forall z_1 \dots \forall z_6\ f(t; x_1, x_2; y_1, \dots, y_7; z_1, \dots, z_6) = 0$$

for some $f \in \mathbb{Z}[T; X_1, X_2; Y_1, \dots, Y_7; Z_1, \dots, Z_6]$, i.e., for any $t \in \mathbb{Q}$,

$$t \in \mathbb{Z} \text{ iff } \phi(t) \text{ holds in } \mathbb{Q}.$$

In 2009, Bjorn Poonen ([P09a]) managed to find an $\forall\exists$-definition with 2 universal and 7 existential quantifiers (earlier, in Cornelissen and Zahidi [2007], an $\forall\exists$-definition with just one universal quantifier was proved modulo an open conjecture on elliptic curves).

In Koenigsmann [2016], the author then provided a $\forall$-definition of $\mathbb{Z}$ in $\mathbb{Q}$:

Theorem 3.1. *There is a polynomial $g \in \mathbb{Z}[T; X_1, \dots, X_{418}]$ such that, for all $t \in \mathbb{Q}$,*

$$t \in \mathbb{Z} \textit{ iff } \forall \overline{x} \in \mathbb{Q}^{418}\ g(t; \overline{x}) \neq 0.$$

If one measures logical complexity in terms of the number of changes of quantifiers then this is the simplest definition of $\mathbb{Z}$ in $\mathbb{Q}$, and, in fact, it is the simplest possible: there is no quantifier-free definition of $\mathbb{Z}$ in $\mathbb{Q}$.

Corollary 3.2. $\mathbb{Q} \setminus \mathbb{Z}$ *is diophantine in* $\mathbb{Q}$.

Corollary 3.3. $Th_{\forall\exists}(\mathbb{Q})$ *is undecidable.*

Theorem 3.1 came somewhat unexpected because it does not give what you would like to have, namely an existential definition of $\mathbb{Z}$ in $\mathbb{Q}$. However, if you had the latter the former would follow:

Observation 3.4. *If there is an existential definition of $\mathbb{Z}$ in $\mathbb{Q}$ then there is also a universal one.*

Proof: If $\mathbb{Z}$ is diophantine in $\mathbb{Q}$ then so is

$$\mathbb{Q} \setminus \mathbb{Z} = \{x \in \mathbb{Q} \mid \exists m, n, a, b \in \mathbb{Z} \text{ with } n \neq 0, \pm 1,\ am + bn = 1 \text{ and } m = xn\}$$

□

The machinery for proving these three first-order definitions of $\mathbb{Z}$ in $\mathbb{Q}$ is not very heavy: Julia Robinson made essentially use of the Hasse-Minkowski Local-Global Principle for quadratic forms, Bjorn Poonen augmented that using the Hasse bound for the number of rational points on genus-1 curves over finite fields (and he ingeniously rearranged the use of quadratic form theory), while in Koenigsmann [2016] the Quadratic Reciprocity Law came in as additional tool, and then some elementary tricks (inspired by the model theory of valued fields) for transforming existential formulas into universal ones were needed to complete the proof.

Using more serious number theory, (Park [2013])has generalised Theorem 3.1 to number fields:

Theorem 3.5. *For any number field K, the ring of integers $\mathcal{O}_K$ is universally definable in K.*

In the course of the proof of Koenigsmann [2016] many new diophantine subsets of $\mathbb{Q}$ emerged, for example the set of non-squares turned out to be diophantine (this was obtained earlier in Poonen [2009b] using much deeper techniques). If, however, $\mathbb{Z}$ was also diophantine in $\mathbb{Q}$ then there would be many more important diophantine subsets of $\mathbb{Q}$, for example the set of tuples of coefficients of irreducible polynomials (of fixed degree) over $\mathbb{Q}$. Later, Philip Dittmann managed to prove this unconditionally and in much greater generality (Dittmann [2016]):

Theorem 3.6. *Irreducibility of polynomials over global fields is diophantine.*

3.2 Hilbert's 10th Problem for number rings using elliptic curves. In this section only one major achievement is being reported on. There is a multitude of surveys on the subject, each with its own emphasis. For the interested reader, let us mention at least some of them: R. M. Robinson [1951], Mazur [1994], Pheidas [1994], Y. Matiyasevich [2000], Pheidas and Zahidi [2000], Shlapentokh [2000], Poonen [2003], Shlapentokh [2007], Poonen [2008], and Koenigsmann [2014].

For number rings and number fields, the question of decidability has been answered in the negative by J. Robinson [1959]. The question whether Hilbert's 10th Problem is solvable is much harder. Given that we don't know the answer over $\mathbb{Q}$ (though almost everyone working in the field believes it to be no) there is even less hope that we find the answer for arbitrary number fields in the near future. For number *rings* the situation is much better.

Let K be a number field with ring of integers $\mathcal{O}_K$. Then Hilbert's 10th Problem could be shown to be unsolvable over $\mathcal{O}_K$ in the following cases:

- if K is totally real or a quadratic extension of a totally real number field (Denef [1975], Denef and Lipshitz [1978] and Denef [1980]),
- if $[K : \mathbb{Q}] \geq 3$ and $c_K = 2$ (Pheidas [1988])[1],
- if $K/\mathbb{Q}$ is abelian (Shapiro and Shlapentokh [1989]).

In each of the proofs the authors managed to find an existential definition of $\mathbb{Z}$ in $\mathcal{O}_K$ using Pell-equations, the Hasse-Minkowski Local-Global Principle (which holds in all number fields) and ad hoc methods that are very specific to each of these special cases.

The hope for a uniform proof of the existential undecidability of all number rings only emerged when elliptic curves were brought into the game:

Theorem 3.7 ([Poo02]). *Let K be a number field. Assume*[2] *there is an elliptic curve E over $\mathbb{Q}$ with $\mathtt{rk}(E(\mathbb{Q})) = \mathtt{rk}(E(K)) = 1$. Then $\mathbb{Z}$ is existentially definable in $\mathcal{O}_K$ and so Hilbert's 10th Problem over $\mathcal{O}_K$ is unsolvable.*

In his proof, Poonen uses divisibility relations for denominators of x-coordinates of $n \cdot P$, where $P \in E(K) \setminus E_{tor}(K)$ and $n \cdot P \in E(\mathbb{Q})$ (for a similar approach cf. Cornelissen, Pheidas, and Zahidi [2005]).

The assumption made in the theorem turns out to hold modulo a generally believed conjecture, the so called *Tate-Shafarevich Conjecture*. For an elliptic curve E over a number field K, it refers to the *Tate-Shafarevich group* (or *Shafarevich-Tate group*) $Ш_{E/K}$, an abelian group defined via cohomology groups. It measures the deviation from a local-global principle for rational points on E.

---

[1] c_K denotes the *class number* of K, that is, the size of the ideal class group of K. It measures how far $\mathcal{O}_K$ is from being a PID: $c_K = 1$ iff $\mathcal{O}_K$ is a PID, so $c_K = 2$ is 'the next best'. It is not known whether there are infinitely many number fields with $c_K = 1$.

[2] The set $E(K)$ of K-rational points of E is a finitely generated abelian group isomorphic to the direct product of its torsion subgroup $E_{tor}(K)$ and a free abelian group of rank '$\mathtt{rk}(E(K))$'.

Tate-Shafarevich Conjecture *$Ш_{E/K}$ is finite.*

Weak Tate-Shafarevich Conjecture $\dim_{\mathbb{F}_2} Ш_{E/K}/2$ *is even.*

The latter follows from the former due to the Cassels pairing (Theorem 4.14 in Silverman [1986] which is an excellent reference on elliptic curves).

Theorem 3.8 ([MR10]). *Let K be a number field. Assume the weak Tate-Shafarevich Conjecture for all elliptic curves E/K. Then there is an elliptic curve $E/\mathbb{Q}$ with $\mathtt{rk}(E(\mathbb{Q})) = \mathtt{rk}(E(K)) = 1$.*

Taking those two theorems together you obtain immediately the following

Corollary 3.9. *Let K be a number field. Assume the weak Tate-Shafarevich Conjecture for all elliptic curves E/K. Then Hilbert's 10th Problem is unsolvable over $\mathcal{O}_K$.*

3.3 Global fields of positive characteristic. It is natural to ask decidability questions not only over number fields, but also over global fields of positive characteristic, i.e., algebraic function fields in one variable over finite fields, and also, more generally, for function fields.

Hilbert's 10th Problem (with t resp. t_1, t_2 in the language) has been shown to be unsolvable for the following function fields:

- $\mathbb{R}(t)$ (Denef [1978]),
- $\mathbb{C}(t_1, t_2)$ (Kim and Roush [1992]),
- $\mathbb{F}_q(t)$ (Pheidas [1991] and Videla [1994]),
- finite extensions of $\mathbb{F}_q(t)$ (Shlapentokh [1992] and Eisenträger [2003]).

The first two cases were achieved by existentially defining $\mathbb{Z}$ in the field, and then applying Matiyasevich's Theorem. This is, clearly, not possible in the last two cases. Instead of existentially *defining* $\mathbb{Z}$ the authors existentially *interpret* $\mathbb{Z}$ via elliptic curves: the multiplication by n-map on an elliptic curve E/K where $E(K)$ contains non-torsion points easily gives a diophantine interpretation of the additive group $\langle \mathbb{Z}; + \rangle$. The difficulty is to find an elliptic curve E/K such that there is also an existential definition for multiplication on that additive group.

For the ring of *polynomials* $\mathbb{F}_q[t]$, Demeyer has even shown the analogue of the DPRM-Theorem: listible subsets are diophantine (Demeyer [2007]).

Generalizing earlier results (Cherlin [1984], Duret [1986] and Pheidas [2004]), it is shown in Eisenträger and Shlapentokh [2009], that the *full* first-order theory of any function field of characteristic > 2 is undecidable.

For analogues of Hilbert's 10th Problem for fields of meromorphic or analytic functions cf., e.g., Rubel [1995], Vidaux [2003] and Pasten [2013].

4 Two infinite extensions of $\mathbb{Q}_p$

Let us recall that the field $\mathbb{Q}_p$ is axiomatized as a valued field (K, v) satisfying the following four axioms:

(1) (K, v) is henselian of mixed characteristic $(0, p)$,

(2)' $Kv = \mathbb{F}_p$,

(3)' $vK \equiv \mathbb{Z}$, so vK is a $\mathbb{Z}$-group,

(4)' $v(p)$ is minimal positive.

It is an immediate consequence of the main result of Derakhshan and Macintyre [2016] that the field $\mathbb{Q}_p^{ur}$, the maximal unramified extension of $\mathbb{Q}_p$ (obtained from $\mathbb{Q}_p$ by adjoining all prime to p roots of unity), is model complete, that is, every first order definable subset is already existentially definable. Using this, you can easily give an axiomatization of $\mathbb{Q}_p^{ur}$, namely as valued field (K, v) satisfying these axioms:

(1) (K, v) is henselian of mixed characteristic $(0, p)$,

(2) $Kv = \overline{Kv}$, so the residue field is algebraically closed,

(3)' $vK \equiv \mathbb{Z}$, so vK is a $\mathbb{Z}$-group,

(4)' $v(p)$ is minimal positive.

The next natural challenge is to find an axiomatization for $\mathbb{Q}_p^{ab}$, the maximal abelian extension of $\mathbb{Q}_p$, which, by the local Kronecker-Weber Theorem, is obtained from $\mathbb{Q}_p$ by adjoining all roots of unity. The axiomatization suggested (but not yet proved to be complete) in Koenigsmann [2018] is the axiomatization as valued field (K, v) satisfying these axioms:

(1) (K, v) is henselian of mixed characteristic $(0, p)$,

(2) $Kv = \overline{Kv}$,

(3) $vK \equiv \frac{1}{p^\infty}\mathbb{Z}$,

(4) $q \nmid v(1 - \zeta_p)$ for any prime $q \neq p$,

(5) $K \cap \overline{\mathbb{Q}} = \mathbb{Q}_p^{ab} \cap \overline{\mathbb{Q}}$,

(6) $v = v_K^p$,

(7) the Frobenius map $x \mapsto x^p$ is surjective on $\mathcal{O}_v / p\mathcal{O}_v$.

Here $\frac{1}{p^\infty}\mathbb{Z}$ is the ordered subgroup of the group of rational numbers having only p-th powers as denominators, ζ_p is a primitive p-th root of unity, and v_K^p is the *canonical p-henselian valuation* on K, that is here the coarsest p-henselian valuation with p-closed residue field, where p-henselian means that the valulation extends uniquely to every Galois extension of degree p. That these axioms can be expressed by (recursive sets of) first-order formulas is not too hard to show, except for axiom (6), for which this is proved in Jahnke and Koenigsmann [2015]. It is also not too difficult to check that all these axioms are true in $\mathbb{Q}_p^{ab}$. However, it requires substantial work to prove that these axioms are independent, that is, for each of the seven axioms one finds a valued field not satisfying this particular axiom, but satisfying all the other axioms (this is done in Koenigsmann [2018]. The planned strategy for establishing that these axioms are complete is via showing quantifier elimination in a variant of the Macintyre language for valued fields including n-th power predicates.

References

Sylvy Anscombe and Arno Fehm (2016). "The existential theory of equicharacteristic henselian valued fields". *Algebra Number Theory* 10.3, pp. 665–683. MR: 3513134 (cit. on p. 67).

Sylvy Anscombe and Franz-Viktor Kuhlmann (2016). "Notes on extremal and tame valued fields". *J. Symb. Log.* 81.2, pp. 400–416. MR: 3519438 (cit. on p. 68).

Will Anscombe and Jochen Koenigsmann (2014). "An existential $\varnothing$-definition of $\mathbb{F}_q[[t]]$ in $\mathbb{F}_q((t))$". *J. Symb. Log.* 79.4, pp. 1336–1343. MR: 3343541 (cit. on p. 68).

James Ax and Simon Kochen (1965). "Diophantine problems over local fields. I". *Amer. J. Math.* 87, pp. 605–630. MR: 0184930 (cit. on p. 66).

Salih Azgin, Franz-Viktor Kuhlmann, and Florian Pop (2012). "Characterization of extremal valued fields". *Proc. Amer. Math. Soc.* 140.5, pp. 1535–1547. MR: 2869138 (cit. on p. 68).

Gregory L. Cherlin (1984). "Undecidability of rational function fields in nonzero characteristic". In: *Logic colloquium '82 (Florence, 1982)*. Vol. 112. Stud. Logic Found. Math. North-Holland, Amsterdam, pp. 85–95. MR: 762105 (cit. on p. 72).

Raf Cluckers, Jamshid Derakhshan, Eva Leenknegt, and Angus Macintyre (2013). "Uniformly defining valuation rings in Henselian valued fields with finite or pseudo-finite residue fields". *Ann. Pure Appl. Logic* 164.12, pp. 1236–1246. MR: 3093389 (cit. on p. 68).

Gunther Cornelissen, Thanases Pheidas, and Karim Zahidi (2005). "Division-ample sets and the Diophantine problem for rings of integers". *J. Théor. Nombres Bordeaux* 17.3, pp. 727–735. MR: 2212121 (cit. on p. 71).

Gunther Cornelissen and Karim Zahidi (2007). "Elliptic divisibility sequences and undecidable problems about rational points". *J. Reine Angew. Math.* 613, pp. 1–33. MR: 2377127 (cit. on p. 69).

Martin Davis, Hilary Putnam, and Julia Robinson (1961). "The decision problem for exponential diophantine equations". *Ann. of Math. (2)* 74, pp. 425–436. MR: 0133227 (cit. on p. 64).

Jeroen Demeyer (2007). "Recursively enumerable sets of polynomials over a finite field are Diophantine". *Invent. Math.* 170.3, pp. 655–670. MR: 2357505 (cit. on p. 72).

J. Denef (1975). "Hilbert's tenth problem for quadratic rings". *Proc. Amer. Math. Soc.* 48, pp. 214–220. MR: 0360513 (cit. on p. 71).

– (1978). "The Diophantine problem for polynomial rings and fields of rational functions". *Trans. Amer. Math. Soc.* **242**, pp. 391–399. MR: 0491583 (cit. on p. 72).

– (1980). "Diophantine sets over algebraic integer rings. II". *Trans. Amer. Math. Soc.* 257.1, pp. 227–236. MR: 549163 (cit. on p. 71).

J. Denef and L. Lipshitz (1978). "Diophantine sets over some rings of algebraic integers". *J. London Math. Soc. (2)* 18.3, pp. 385–391. MR: 518221 (cit. on p. 71).

Jan Denef, Leonard Lipshitz, Thanases Pheidas, and Jan Van Geel, eds. (2000). *Hilbert's tenth problem: relations with arithmetic and algebraic geometry*. Vol. 270. Contemporary Mathematics. Papers from the workshop held at Ghent University, Ghent, November 2–5, 1999. American Mathematical Society, Providence, RI, pp. xii+367. MR: 1802007.

Jan Denef and Hans Schoutens (2003). "On the decidability of the existential theory of $\mathbb{F}_p[\![t]\!]$". In: *Valuation theory and its applications, Vol. II (Saskatoon, SK, 1999)*. Vol. 33. Fields Inst. Commun. Amer. Math. Soc., Providence, RI, pp. 43–60. MR: 2018551 (cit. on p. 67).

Jamshid Derakhshan and Angus Macintyre (Mar. 2016). "Model Completeness for Henselian Fields with finite ramification valued in a Z-Group". arXiv: 1603.08598 (cit. on p. 73).

Philip Dittmann (Jan. 2016). "Irreducibility of Polynomials over Global Fields is Diophantine". arXiv: 1601.07829 (cit. on p. 70).

Jean-Louis Duret (1986). "Sur la théorie élémentaire des corps de fonctions". *J. Symbolic Logic* 51.4, pp. 948–956. MR: 865921 (cit. on p. 72).

Kirsten Eisenträger (2003). "Hilbert's tenth problem for algebraic function fields of characteristic 2". *Pacific J. Math.* 210.2, pp. 261–281. MR: 1988534 (cit. on p. 72).

Kirsten Eisenträger and Alexandra Shlapentokh (2009). "Undecidability in function fields of positive characteristic". *Int. Math. Res. Not. IMRN* 21, pp. 4051–4086. MR: 2549950 (cit. on p. 72).

Antonio J. Engler and Alexander Prestel (2005). *Valued fields*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, pp. x+205. MR: 2183496 (cit. on p. 67).

Yu. L. Ershov (2004). "Extremal valued fields". *Algebra Logika* 43.5, pp. 582–588, 631. MR: 2112060 (cit. on p. 68).

– (2009). "$*$-extremal valued fields". *Sibirsk. Mat. Zh.* 50.6, pp. 1280–1284. MR: 2603869 (cit. on p. 68).

Ju. L. Eršov (1965). "On elementary theories of local fields". *Algebra i Logika Sem.* 4.2, pp. 5–30. MR: 0207686 (cit. on p. 66).

Michael Fried and Moshe Jarden (2008). *Field arithmetic*. 3rd ed. Springer (cit. on p. 66).

Kurt Gödel (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". *Monatsh. Math. Phys.* 38.1, pp. 173–198. MR: 1549910 (cit. on p. 65).

David Hilbert (1900). "Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker Kongress zu Paris 1900". In: *Nachr. K. Ges. Wiss., Göttingen, Math.-Phys. Kl.* Pp. 253–297 (cit. on p. 63).

Franziska Jahnke and Jochen Koenigsmann (2015). "Uniformly defining p-henselian valuations". *Ann. Pure Appl. Logic* 166.7-8, pp. 741–754. MR: 3344577 (cit. on p. 74).

K. H. Kim and F. W. Roush (1992). "Diophantine undecidability of $\mathbf{C}(t_1, t_2)$". *J. Algebra* 150.1, pp. 35–44. MR: 1174886 (cit. on p. 72).

Jochen Koenigsmann (2014). "Undecidability in number theory". In: *Model theory in algebra, analysis and arithmetic*. Vol. 2111. Lecture Notes in Math. Springer, Heidelberg, pp. 159–195. MR: 3330199 (cit. on pp. 70, 71).

– (2016). "Defining $\mathbb{Z}$ in $\mathbb{Q}$". *Ann. of Math. (2)* 183.1, pp. 73–93. MR: 3432581 (cit. on pp. 69, 70).

– (2018). "On the decidability of $\mathbb{Q}_p^{ab}$". To appear (cit. on pp. 73, 74).

Yuri Matiyasevich (2000). "Hilbert's tenth problem: what was done and what is to be done". In: *Hilbert's tenth problem: relations with arithmetic and algebraic geometry (Ghent, 1999)*. Vol. 270. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 1–47. MR: 1802008 (cit. on p. 70).

Yuri V. Matiyasevich (1970a). "Diofantovost' perechislimykh mnozhestv". *Soviet Math. Doklady* 11. Translated of the Russian original Y. V. Matiyasevich [1970b], pp. 354–358 (cit. on p. 76).

– (1970b). "Diofantovost' perechislimykh mnozhestv". *Dokl. AN SSSR* 191 (2). Translated to English in Y. V. Matiyasevich [1970a], pp. 278–282 (cit. on pp. 64, 76).

B. Mazur (1994). "Questions of decidability and undecidability in number theory". *J. Symbolic Logic* 59.2, pp. 353–371. MR: 1276620 (cit. on p. 70).

B. Mazur and K. Rubin (2010). "Ranks of twists of elliptic curves and Hilbert's tenth problem". *Invent. Math.* 181.3, pp. 541–575. MR: 2660452.

Jennifer Park (2013). "A universal first-order formula defining the ring of integers in a number field". *Math. Res. Lett.* 20.5, pp. 961–980. MR: 3207365 (cit. on p. 70).

Hector Pasten (2013). "Powerful values of polynomials and a conjecture of Vojta". *J. Number Theory* 133.9, pp. 2964–2998. MR: 3057059 (cit. on p. 72).

Thanases Pheidas (1988). "Hilbert's tenth problem for a class of rings of algebraic integers". *Proc. Amer. Math. Soc.* 104.2, pp. 611–620. MR: 962837 (cit. on p. 71).

– (1991). "Hilbert's tenth problem for fields of rational functions over finite fields". *Invent. Math.* 103.1, pp. 1–8. MR: 1079837 (cit. on p. 72).

– (1994). "Extensions of Hilbert's 10th Problem". *J. Symb. Logic* 59.2, pp. 372–397 (cit. on p. 70).

– (2004). "Endomorphisms of elliptic curves and undecidability in function fields of positive characteristic". *J. Algebra* 273.1, pp. 395–411. MR: 2032468 (cit. on p. 72).

Thanases Pheidas and Karim Zahidi (2000). "Undecidability of existential theories of rings and fields: a survey". In: *Hilbert's tenth problem: relations with arithmetic and algebraic geometry (Ghent, 1999)*. Vol. 270. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 49–105. MR: 1802009 (cit. on p. 70).

Bjorn Poonen (2002). "Using elliptic curves of rank one towards the undecidability of Hilbert's tenth problem over rings of algebraic integers". In: *Algorithmic number theory (Sydney, 2002)*. Vol. 2369. Lecture Notes in Comput. Sci. Springer, Berlin, pp. 33–42. MR: 2041072.

– (2003). "Hilbert's 10th problem over rings of number-theoretic interest" (cit. on pp. 70, 71).

– (2008). "Undecidability in number theory". *Notices Amer. Math. Soc.* 55.3, pp. 344–350. MR: 2382821 (cit. on pp. 70, 71).

– (2009a). "Characterizing integers among rational numbers with a universal-existential formula". *Amer. J. Math.* 131.3, pp. 675–682. MR: 2530851.

– (2009b). "The set of nonsquares in a number field is Diophantine". *Math. Res. Lett.* 16.1, pp. 165–170. MR: 2480570 (cit. on p. 70).

Florian Pop (1996). "Embedding problems over large fields". *Ann. of Math. (2)* 144.1, pp. 1–34. MR: 1405941 (cit. on p. 68).

Alexander Prestel and Peter Roquette (1984). *Formally p-adic fields*. Vol. 1050. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. v+167. MR: 738076 (cit. on p. 66).

Julia Robinson (1949). "Definability and decision problems in arithmetic". *J. Symbolic Logic* 14, pp. 98–114. MR: 0031446 (cit. on pp. 65, 69).

– (1959). "The undecidability of algebraic rings and fields". *Proc. Amer. Math. Soc.* 10, pp. 950–957. MR: 0112842 (cit. on p. 71).

Raphael M. Robinson (1951). "Undecidable rings". *Trans. Amer. Math. Soc.* 70, pp. 137–159. MR: 0041081 (cit. on p. 70).

Lee A. Rubel (1995). "An essay on Diophantine equations for analytic functions". *Exposition. Math.* 13.1, pp. 81–92. MR: 1339814 (cit. on p. 72).

Harold N. Shapiro and Alexandra Shlapentokh (1989). "Diophantine relationships between algebraic number fields". *Comm. Pure Appl. Math.* 42.8, pp. 1113–1122. MR: 1029120 (cit. on p. 71).

Alexandra Shlapentokh (1992). "Hilbert's tenth problem for rings of algebraic functions in one variable over fields of constants of positive characteristic". *Trans. Amer. Math. Soc.* 333.1, pp. 275–298. MR: 1091233 (cit. on p. 72).

– (2000). "Hilbert's tenth problem over number fields, a survey". In: *Hilbert's tenth problem: relations with arithmetic and algebraic geometry (Ghent, 1999)*. Vol. 270. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 107–137. MR: 1802010 (cit. on pp. 70, 71).

– (2007). *Hilbert's tenth problem*. Vol. 7. New Mathematical Monographs. Diophantine classes and extensions to global fields. Cambridge University Press, Cambridge, pp. xiv+320. MR: 2297245 (cit. on pp. 70, 71).

Joseph H. Silverman (1986). *The arithmetic of elliptic curves*. Vol. 106. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xii+400. MR: 817210 (cit. on p. 72).

X. Vidaux (2003). "An analogue of Hilbert's 10th problem for fields of meromorphic functions over non-Archimedean valued fields". *J. Number Theory* 101.1, pp. 48–73. MR: 1979652 (cit. on p. 72).

Carlos R. Videla (1994). "Hilbert's tenth problem for rational function fields in characteristic 2". *Proc. Amer. Math. Soc.* 120.1, pp. 249–253. MR: 1159179 (cit. on p. 72).

Received 2018-03-11.

JOCHEN KOENIGSMANN
MATHEMATICAL INSTITUTE
RADCLIFF OBSERVATORY QUARTER
OXFORD OX2 6GG
UK
koenigsmann@maths.ox.ac.uk

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (79–100)

PROOF-THEORETIC METHODS IN NONLINEAR ANALYSIS

ULRICH KOHLENBACH

Abstract

We discuss applications of methods from proof theory, so-called proof interpretations, for the extraction of explicit bounds in convex optimization, fixed point theory, ergodic theory and nonlinear semigroup theory.

1 Introduction: Proof Theory, Hilbert's Program and Kreisel's 'Unwinding of Proofs'

Proof theory has its origin in what has been called 'Hilbert's program': Since the 19th century noneffective and nonfinitary (set-theoretic) principles became increasingly important which raised the issue of their legitimacy. Hilbert's approach was to establish the consistency of a suitable formalization T of mathematics (first number theory and then analysis and set theory) within some finitary reasoning T_{fin}. In the language of number theory and with a minimal amount of number-theoretic tools one can express the consistency of T (axiomatized by an effective list of axioms) as a purely universal number-theoretic sentence (a so-called Π^0_1-sentence)

$$Con_T := \forall n \in \mathbb{N} \neg Prov_T(n, \ulcorner 0 = 1 \urcorner)$$

which states that no $n \in \mathbb{N} := \{0, 1, 2, \ldots\}$ is the code of a T-proof of $0 = 1$.

Consider now an arbitrary Π^0_1-sentence (called a 'real statement' by Hilbert) $S := \forall n \in \mathbb{N}\ (t(n) = 0)$, where t is some primitive recursive function term. If S is provable in T (using any nonfinitary 'ideal elements' of T), then also $T_{fin} + Con_T$ proves S (see Smorynski [1977][5.2.1]). So if Con_T could be proved in T_{fin}, one could convert the 'ideal' proof of S in T into a finitistic proof of S in T_{fin}.

Obviously, Gödel's second incompleteness theorem rules out that the consistency of any

MSC2010: primary 03F10; secondary 03F35, 47H10, 47H20, 47H25, 90C30.
Keywords: Proof mining, effective bounds, convex optimization, fixed points, ergodic theory, nonlinear semigroups.

$T \supseteq T_{fin}$ can be established inside of T_{fin}. Nevertheless, Hilbert's program gave rise to many 'relative consistency proofs' where the consistency of T is reduced to that of an in **some** sense more elementary theory T'. 'More elementary' often is related to being 'quantifier-free': e.g. Gentzen's proof-theoretic analysis of first-order number theory PA (Gentzen [1936]) reduces logically complex instances of ordinary induction to quantifier-free instances of transfinite induction (along ordinals $\alpha < \varepsilon_0$) and Gödel's consistency proof for PA (Gödel [1958]) reduces PA to a quantifier-free calculus of so-called primitive recursive functionals of finite type (considered already by Hilbert himself in Hilbert [1926]).
In the early 50's, Georg Kreisel suggested to re-orient proof theory by applying proof-theoretic methods - which in some way eliminate quantifiers in terms of quantifier-free constructions - to proofs of theorems which are not purely universal (as consistency statements) but e.g. of the form

$$(*)\ \forall n \in \mathbb{N}\, \exists m \in \mathbb{N}\, A_{qf}(n,m) \quad (A_{qf} \text{ quantifier-free}).$$

Kreisel noted that the respective consistency proofs for PA due to Gentzen (see Kreisel [1951, 1952]) and Gödel resp. (see Kreisel [1959] (3.4)) actually characterize the class of subrecursive functions f needed to realize $(*)$ in the form

$$\forall n \in \mathbb{N}\, A_{qf}(n, f(n))$$

for theorems $(*)$ which are provable in PA, namely as the class of $\alpha < \varepsilon_0$-recursive functions (in the case of Gentzen's proof) and - equivalently - as the class of functions definable in the aforementioned calculus of primitive recursive functionals (in the case of Gödel's proof Gödel [1941, 1958]), see also Parsons [1972].
While such results concern (the provability of $\forall\exists$-sentences in) formal systems such as PA rather than individual proofs, Kreisel already in Kreisel [1952] also launched the program of analyzing specific prima facie nonconstructive proofs with the aim of extracting new (e.g. effective) information on the theorem proven:

Input: A (prima facie) noneffective proof P of a conclusion C.

Goal: Additional information on C such as:

- effective bounds,
- algorithms,
- continuous dependency or full independence from certain parameters,
- generalizations of proofs: weakening of premises.

Kreisel's examples and suggestions for applications mainly concerned proofs in number theory. E.g. in Kreisel [1982], Kreisel suggested to analyze finiteness statements such as Roth's theorem in diophantine approximation with the aim of extracting bounds on the number of solutions. In Luckhardt [1989], Luckhardt extracted the first polynomial such bound for Roth's theorem from a proof due to Esnault and Viehweg (independently, this result was also obtained in Bombieri and van der Poorten [1988]). Since the 90's, the program has been developed most systematically and with specially designed so-called logical metatheorems (see the next section) in the context of nonlinear analysis ('proof mining'). Also while Kreisel's unwindings were based on techniques related to cut-elimination (Herbrand theory, ε-substitution etc.) the applications to analysis are all based on functional interpretations which have their origin in Gödel's 'Dialectica' interpretation on which Gödel's aforementioned consistency proof is based.

2 Logical metatheorems for bound extractions

In order to establish general theorems on the extractability of effective uniform bounds from given proofs one has to set up an appropriate formal deductive context. As the bound extraction methods are based on modern ('monotone') extensions and variants (see Kohlenbach [2008a]) of Gödel's functional interpretation (Gödel [1941, 1958]) one uses formal systems formulated in the language of functionals in all finite types such as appropriate forms of Peano arithmetic in all finite types PA^ω. In such systems one already can represent complete separable metric ('Polish') spaces (X, d) as continuous images of the Baire space $\mathbb{N}^{\mathbb{N}}$. However, this requires the separability of the space X and for separable spaces one can show that the independence of the extracted bounds from parameters in subspaces of X in general can only be expected if these subspaces are compact (see Kohlenbach [2008a] for discussions of this point). Many theorems in nonlinear analysis, however, involve - in addition to **concrete** Polish spaces such as $\mathbb{R}$ - general classes of **abstract** spaces X (e.g. general Hilbert spaces) which are not required to be separable and one can extract bounds that are independent from parameters in X (and even functions $T : X \to X$) if general metric bounds ('majorants') are given.

Many abstract types of metric structures can be added as atoms to our formal systems. E.g. this applies to metric, W-hyperbolic (see below), CAT(0), CAT(1), δ-hyperbolic, normed, uniformly convex, Hilbert, abstract L^p, abstract $C(K)$ spaces and $\mathbb{R}$-trees, and, in fact, all normed structures that are axiomatizable in so-called positive bounded logic (see Günzel and Kohlenbach [2016]). In order to be able to speak about such spaces one adds a **new base type** X to the formal system and forms all finite types over $\mathbb{N}, X$ (see Kohlenbach [2005b]; one may also have several such types: see Kohlenbach [2008a], section 17.6). One also adds constants for the metric d_X or normed space operators with appropriate axioms that characterize the class of structures in question.

Condition: the defining axioms must have a monotone functional interpretation (possibly with the addition of appropriate moduli, see Kohlenbach [2008a]).

Counterexamples (to the extractability of uniform bounds) exist for the classes of strictly convex or separable spaces which get upgraded by the monotone functional interpretation to uniformly convex resp. boundedly compact spaces.

Formal systems for analysis with abstract spaces X

Types: (i) $\mathbb{N}, X$ are types, (ii) with ρ, τ also $\rho \to \tau$ is a type.

Functionals of type $\rho \to \tau$ map type-ρ objects to type-τ objects.

$\mathrm{PA}^{\omega,X}$ is the extension of PA to all types, $\mathcal{A}^{\omega,X}$:=$\mathrm{PA}^{\omega,X}$+DC, where

DC: axiom schema of dependent choice for all types,

which implies the axiom schema of countable choice and so, applied to the law-of-excluded middle, full comprehension for numbers

$$\text{CA: } \exists f^{\mathbb{N}\to\mathbb{N}}\ \forall n^{\mathbb{N}}\ (f(n) = 0 \leftrightarrow A(n)),$$

where $A(n)$ may contain quantifiers (and parameters) of arbitrary types.

$\mathcal{A}^{\omega}[X, d, \ldots]$ results by adding constants $d_X, \ldots$ with axioms expressing that $(X, d, \ldots)$ is a nonempty metric, hyperbolic …space (we deviate here from the notation used in Kohlenbach [ibid.] where this theory is denoted by $\mathcal{A}^{\omega}[X, d, \ldots]_{-b}$, and $\mathcal{A}^{\omega}[X, d, \ldots]$ denotes the theory with an axiom stating the boundedness of (X, d) by some constant b being added).

A warning concerning equality: our formal theories only have a **quantifier-free rule of extensionality** (with A_{qf} being a quantifier-free formula)

$$\frac{A_{qf} \to s =_{\rho} t}{A_{qf} \to r[s/x] =_{\tau} r[t/x]},$$

where only $x =_{\mathbb{N}} y$ is a primitive predicate but for X and $\rho \to \tau$ one defines

$$x^X =_X y^X :\equiv d_X(x, y) =_{\mathbb{R}} 0_{\mathbb{R}},\ x =_{\rho\to\tau} y :\equiv \forall v^{\rho}(x(v) =_{\tau} y(v)).$$

This is crucial as the uniform quantitative rendering of the extensionality axiom $x =_X y \to Tx =_X Ty$ for T of type $X \to X$ implies the uniform continuity of T (on bounded subsets) and we want (in contrast to the setting of current continuous model theory; see, however, the recent Cho [2016]) also to be able to treat discontinuous situations (see Kohlenbach [2008a] for extensive discussions of this point).

Extension of majorizability to the new types: A crucial notion used is an extension of Howard's (Howard [1973]) concept of majorizability to the new types, where we 'bound' an element in a metric space by the distance it has from a fixed reference point $a \in X$ (where $a = 0_X$ in the normed case): let y, x be functionals be of types $\rho, \widehat{\rho} := \rho[\mathbb{N}/X]$ and a^X of type X:

$$x^{\mathbb{N}} \gtrsim^a_{\mathbb{N}} y^{\mathbb{N}} :\equiv x \geq y, \; x^{\mathbb{N}} \gtrsim^a_X y^X :\equiv x \geq d(y, a).$$

For complex types $\rho \to \tau$ this is extended in a hereditary fashion.

Example: for monotone T^* one defines

$$T^* \gtrsim^a_{X \to X} T \equiv \forall n \in \mathbb{N}, x \in X[n \geq d(a, x) \to T^*(n) \geq d(a, T(x))]$$

(see Gerhardy and Kohlenbach [2008] and Kohlenbach [2008a]).
$T : X \to X$ is nonexpansive (n.e.) if $d(T(x), T(y)) \leq d(x, y)$.
Then $\lambda n.n + b \gtrsim^a_{X \to X} T$, if $d(a, T(a)) \leq b$.
Proof mining exhibits the finitary combinatorial kernel of a proof and as a consequence of this it often is easy to generalize things from a normed linear setting to some geodesic setting. In fact, the approach has been particularly useful in the context of hyperbolic spaces which is a variant of notions considered by Takahashi [1970], Goebel and Kirk [1983] and Kirk [1981/82] and Reich and Shafrir [1990] (see Kohlenbach [2005b] for the precise relationship):

Definition 2.1 (Kohlenbach [ibid.]). *A $(W$-$)$hyperbolic space is a triple (X, d, W) where (X, d) is a metric space and $W : X \times X \times [0, 1] \to X$ s.t. for all $x, y, z \in W$ and $\lambda, \tilde{\lambda} \in [0, 1]$*

(i) $d(z, W(x, y, \lambda)) \leq (1 - \lambda)d(z, x) + \lambda d(z, y)$,

(ii) $d(W(x, y, \lambda), W(x, y, \tilde{\lambda})) = |\lambda - \tilde{\lambda}| \cdot d(x, y)$,

(iii) $W(x, y, \lambda) = W(y, x, 1 - \lambda)$,

(iv) $d(W(x, z, \lambda), W(y, w, \lambda)) \leq (1 - \lambda)d(x, y) + \lambda d(z, w)$.

CAT(0)-spaces (Gromov) are hyperbolic spaces (X, d, W) which satisfy the CN-inequality of Bruhat-Tits (determining W uniquely): for all $x, y_0, y_1, y_2 \in X$

$$\begin{cases} d(y_0, y_1) = \frac{1}{2}d(y_1, y_2) = d(y_0, y_2) \to \\ d(x, y_0)^2 \leq \frac{1}{2}d(x, y_1)^2 + \frac{1}{2}d(x, y_2)^2 - \frac{1}{4}d(y_1, y_2)^2. \end{cases}$$

Small types (over $\mathbb{N}, X$) include: $\mathbb{N}, \; \mathbb{N} \to \mathbb{N}, \; X, \; \mathbb{N} \to X, \; X \to X$.

Theorem 2.2 (Gerhardy and Kohlenbach [2008] and Kohlenbach [2008a]). *Let P, K be Polish resp. compact metric spaces (definable in $\mathcal{A}^\omega$), $A_\exists$*[1] *be an $\exists$-formula and $\underline{\tau}$ be a tuple of small types.*
If $\mathcal{A}^\omega[X, d, W]$ proves

$$\forall x \in P\, \forall y \in K\, \forall \underline{z}^{\underline{\tau}}\, \exists v^{\mathbb{N}}\, A_\exists(x, y, \underline{z}, v),$$

then one can extract a computable $\Phi : \mathbb{N}^{\mathbb{N}} \times \underline{\mathbb{N}}^{(\mathbb{N})} \to \mathbb{N}$ s.t. the following holds in every nonempty hyperbolic space: for all representatives $r_x \in \mathbb{N}^{\mathbb{N}}$ of $x \in P$ and all $\underline{z}^{\underline{\tau}}$ and $\underline{z}^* \in \mathbb{N}^{(\mathbb{N})}$ s.t. $\exists a \in X(\underline{z}^* \gtrsim^a_{\underline{\tau}} \underline{z})$:

$$\forall y \in K\, \exists v \le \Phi(r_x, \underline{z}^*)\, A_\exists(x, y, \underline{z}, v).$$

For the case of bounded hyperbolic spaces, see Kohlenbach [2005b].

As a special case of the above metatheorem one has:

Corollary 2.3 (Gerhardy and Kohlenbach [2008] and Kohlenbach [2008a]). *If $\mathcal{A}^\omega[X, d, W]$ proves*

$$\forall x \in P\, \forall y \in K\, \forall z \in X\, \forall T : X \to X\, \big(T \text{ n.e.} \to \exists v \in \mathbb{N}\, A_\exists\big),$$

then one can extract a computable $\Phi : \mathbb{N}^{\mathbb{N}} \times \mathbb{N} \to \mathbb{N}$ s.t. for all $x \in P, b \in \mathbb{N}$

$$\forall y \in K\, \forall z \in X\, \forall T : X \to X\, \big(T \text{ n.e.} \wedge d_X(z, T(z)) \le b \to \exists v \le \Phi(r_x, b) A_\exists\big)$$

holds in all nonempty hyperbolic spaces (X, d, W).

Similar results hold for the other classes of metric and normed structures listed above. In the normed case, one additionally needs $\|z\| \le b$ as an assumption in the conclusion of the corollary.

Remark 2.4. *Usually, proofs in ordinary mathematics only require a small fragment of $\mathcal{A}^\omega[X, d, \ldots]$ with e.g. the binary ('weak') König's lemma* WKL *instead of* DC *and Σ^0_1-induction only, which guarantees the extractability of primitive recursive (in the sense of Kleene) bounds.* WKL *is equivalent to a sentence of the form $\forall f^{\mathbb{N}\to\mathbb{N}}\, \exists b \le_{\mathbb{N}\to\mathbb{N}} 1 \forall x^{\mathbb{N}}\, A_{qf}(f, b, x)$ and can be added to the system via a Skolem constant B with the purely universal axiom $\forall f, x(Bf \le 1 \wedge A_{qf}(f, Bf, x))$ which is satisfiable in the full set-theoretic model and B is trivially majorized by the constant-1 functional in the extracted bound (see* Kohlenbach [2008a]*).*

[1] There are some mild restrictions on the types of the quantified variables in $A_\exists$.

3 General types of applications

3.1 Asymptotic regularity theorems. Consider a metric space (X, d) and a continuous function $F : X \to \mathbb{R}$. Many problems can be stated in the form of finding a zero $z \in X$ of F. Such problems are often algorithmically approached by setting up some iterative procedure resulting in a sequence (x_n) in X which converges to a zero z of F :

$$(*)\ F(\lim_{n\to\infty} x_n) = F(z) = 0.$$

In this case one, in particular, has that

$$(**)\ F(x_n) \to 0.$$

Quite often, $(**)$ holds under much more general conditions than those needed to ensure the convergence of (x_n) itself. In the case of fixed point problems for mappings $T : X \to X$, i.e. the case where $F(x) := d(x, Tx)$, results of the form $(**)$ are usually referred to as asymptotic regularity statements where this term was originally introduced by Browder and Petryshyn [1966] to refer to the property of T that the sequence $x_n := T^n x$ of Picard iterates satisfies $d(x_n, Tx_n) \to 0$. In many cases (see below) $(d(x_n, Tx_n))_{n\in\mathbb{N}}$ for some iterative process not only converges to 0 but does so in a nonincreasing way. In this situation the asymptotic regularity statement can be equivalently written in the form

$$\forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\ (d(x_n, Tx_n) < 2^{-k}) \in \forall\exists$$

and any upper bound $\Phi(k)$ on '$\exists n$' provides a rate of convergence. This means that one can apply the logical metatheorems mentioned in the previous section to extract effective and highly uniform rates of asymptotic regularity even from prima facie noneffective proofs of asymptotic regularity. In fact, this has been achieved in many instances in the context of nonlinear analysis (see some of the applications below and Kohlenbach [2008b, 2017] for general surveys).

3.2 Strong convergence theorems. Suppose that the theorem to be studied is not about an asymptotic regularity result but about the convergence of the sequence (x_n) itself, e.g. towards a zero of F or a fixed point of T. Already the Cauchy property of (x_n)

$$(+)\ \forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\, \forall i, j \geq n\ (d(x_i, x_j) \leq 2^{-k}) \in \forall\exists\forall$$

has too complicated a logical form to directly apply the logical metatheorems on uniform bound extractions and, in fact, there are already simple cases of computable monotone sequences of rational numbers in $[0, 1]$ which do not have a computable rate of convergence (Specker [1949]).

Roughly speaking, one can distinguish the following situations:
1) The proof of the Cauchy property of (x_n) (or of the convergence of (x_n) to some known element $x \in X$) uses - on top of constructive ('intuitionistic') reasoning - at most the law-of-excluded-middle schema LEM for negated formulas

$$\text{LEM}_\neg : \ \neg A \vee \neg\neg A$$

which, in particular covers the case where A is $\exists$-free (e.g. $A \in \Pi^0_1$) as such formulas are equivalent to their double negation (using the stability of the prime formulas in our formal systems).
Alternatively (but not combined), one may use the so-called Markov principle

$$\text{M} : \ \neg\neg\exists n \in \mathbb{N}\, A_{qf}(n) \to \exists n \in \mathbb{N}\, A_{qf}(n) \ \ (A_{qf} \text{ quantifier-free with parameters})$$

together with the following weak form of LEM (weaker than LEM for Π^0_1-formulas):

$$\text{LLPO} : \ \neg(\exists n \in \mathbb{N}\, A_{qf}(n) \wedge \exists n \in \mathbb{N}\, B_{qf}(n)) \to \forall n \in \mathbb{N}\, \neg A_{qf}(n) \vee \forall n \in \mathbb{N}\, \neg B_{qf}(n),$$

where A_{qf}, B_{qf} are quantifier-free formulas. In both scenarios one can set up logical bound extraction metatheorems, where instead of the purely existential formula $A_\exists$ one may now have an arbitrary formula (see Kohlenbach [2008a]). Since $(+)$ is monotone w.r.t. '$\exists n \in \mathbb{N}$' any upper bound on $n \leq \Phi(k)$ in fact is a Cauchy rate for (x_n) and so one can in these cases extract effective rates of convergence.
2) If the proof of the Cauchy property of (x_n) uses LEM for Σ^0_1-formulas (purely existential formulas for natural numbers) as in the case of the Specker sequences from Specker [1949], then one often has the following dichotomy: either one can show that (x_n) converges to the **unique** zero of F or fixed point of T, or one can use the non-uniqueness of the solution to construct an instance of the Cauchy statement in question which provably does not allow for an effective Cauchy rate.

(i) **Unique existence:** in many cases one can obtain effective rates of convergence (and in fact also with a constructive verification of this fact) for (x_n) if (x_n) converges towards a **unique** zero of F resp. fixed point of T: consider a function $F : X \to \mathbb{R}$ on some metric space (X, d) which has exactly one zero z. The uniqueness part

$$(a)\ \forall x, y \in X\ (F(x) = 0 = F(y) \to x = y)$$

can be written equivalently as

$$(b)\ \forall x, y \in X\ \forall k \in \mathbb{N}\ \exists n \in \mathbb{N}\ (|F(x)|, |F(y)| \leq 2^{-n} \to d(x, y) < 2^{-k}) \in \forall\exists.$$

Then logical metatheorems can be applied to extract from a proof of (a) an effective uniform bound $\Phi(k)$ on '$\exists n \in \mathbb{N}$' in (b), which we called in Kohlenbach [1993] a 'modulus of uniqueness', where $\Phi(k)$ depends on x, y only via general majorizing data and, in particular, is independent of x, y if X is bounded (in the case where X can be treated as an abstract space and, otherwise, if X is compact). Suppose now that we can construct some (bounded) sequence (x_n) of approximate zeros, i.e.

$$(c)\ \forall k \in \mathbb{N}\, \exists n \in \mathbb{N} (|F(x_n)| < 2^{-k}) \in \forall\exists$$

from which we then can extract (using again a logical metatheorem) an effective bound $\Psi(k)$ on '$\exists n \in \mathbb{N}$' in (c), then for $\chi(k) := \Psi(\Phi(k))$ we have

$$\forall k \in \mathbb{N}\, \exists n \leq \chi(k)\, (d(x_n, z) < 2^{-k})$$

and, if we even have that $(|F(x_n)|)_n$ is nonincreasing, it follows that χ is a rate of convergence for $\lim x_n = z$. In Briseid [2009], it is shown that for Picard iterations $x_n = T^n x$ for suitable classes of mappings T the aforementioned logical metatheorems can be used to obtain such rates of convergence even when $(|F(x_n)|)_n$ (for $F(x) := d(x, Tx)$) is not nonincreasing which explains the explicit construction of effective rates of convergence for the classes of asymptotic contractions in the sense of Kirk and of uniformly generalized p-contractive mappings given by Briseid (see Briseid [ibid.] and the literature cited there).

(ii) **Non-unique existence:** when F or T possess many zeros resp. fixed points, one usually can construct computable instances of iterative procedures (x_n) (converging to some zero or fixed point) that do not have a computable rate of convergence. In fact, Neumann [2015] shows that this is the case for the usual iterative schemes used in metric fixed point theory, ergodic theory and convex optimization which even for (firmly) nonexpansive selfmappings $T : [0, 1] \to [0, 1]$ fail to have a computable rate of convergence for simple computable such mappings T. One then has to weaken the goal to what has been called an effective rate of metastability: Noneffectively, (+) is equivalent to

$$(++)\ \forall k \in \mathbb{N}\, \forall g \in \mathbb{N}^{\mathbb{N}}\, \exists n \in \mathbb{N}\, \forall i, j \in [n, n + g(n)]\, (d(x_i, x_j) < 2^{-k}) \in \forall\exists,$$

the so-called Herbrand normal of $(+)$, and a bound $\Phi(k, g)$ on '$\exists n$' is a bound for the Kreisel 'no-counterexample interpretation' (Kreisel [1951, 1952]) of the Cauchy property. Since Tao [2008b] calls an interval $[n, n + g(n)]$ with the property in $(++)$ an interval of 'metastability', we call bounds $\Phi(k, g)$ on '$\exists n$' in $(++)$ rates of metastability. If one additionally knows that (x_n) is converging to a zero of F or a fixed point of T with some rate of metastability then one can actually combine both rates into a common one (formulated here for the case of fixed points), i.e. a bound $\Phi(k, g)$ such that for all $k \in \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$

$$(+++)\ \exists n \leq \Phi(k, g)\, \forall i, j \in [n, n + g(n)]\, (d(x_i, x_j), d(x_i, Tx_i) \leq 2^{-k}).$$

If one has a rate of convergence for $d(x_n, Tx_n) \to 0$, one can even achieve that

$$\exists n \leq \Phi(k,g)\, \forall i,j \in [n, n+g(n)]\, \forall l \geq n\, (d(x_i,x_j), d(x_l,Tx_l) \leq 2^{-k})$$

(see e.g. Kohlenbach, Leuştean, and Nicolae [2018] and Kohlenbach [2016], Rem.2.11). The extraction of explicit bounds Φ on the metastable form of Cauchy or convergence statements is of interest for the following reasons:

a) Disregarding bounded quantifiers, the statement $(+++)$ is purely universal ('real') and captures all the mathematical content of the theorem $\lim x_n = x = Tx$: by a fixed piece of proof it implies back the original convergence theorem: forgetting the bound Φ gives the Herbrand normal form which by recursive comprehension (more precisely QF-AC^{0,0} in the terminology of Kohlenbach [2008a]) and LEM implies the Cauchy property and so by arithmetical comprehension (more precisely Π^0_1-AC^{0,0} in our formal context, see Kohlenbach [ibid.]) the convergence of (x_n). Applying $(+++)$ to the constant function $g(n) := K \in \mathbb{N}$ shows the existence of $i \geq K$ with $d(x_i, Tx_i) < 2^{-k}$ which - together with the continuity of T - gives $Tx = x$ for $x := \lim x_n$.

b) The proof-theoretic extraction of a rate of metastability from a convergence proof exhibits the finitary combinatorial content of that proof which may lead to generalizations of the resulting metastable statement and so - when unpacked into the full convergence statement (see above) - to generalized convergence theorems.

c) The concrete bounds extracted are of numerically interest: often they provide explicit information on the **algorithmic learnability** of a rate of convergence which - if a gap condition is satisfied - yields **oscillation bounds** (Avigad and Rute [2015] and Kohlenbach and Safarik [2014] and Section 5 below).

d) In many cases, asymptotic regularity is just the special case of metastability where $g(n) := 1$, e.g. for Picard iterates of nonexpansive functions T.
Some history:

- 2004, first rate of metastability (for the asymptotic regularity of asymptotically nonexpansive mappings) extracted (Kohlenbach and Lambov [2004]).
- 2005, rate of metastability for Krasnoselski-Mann iterations of nonexpansive selfmappings $T : X \to X$ of compact hyperbolic spaces X (Kohlenbach [2005a]).
- 2007, Tao [2008b] introduced the term 'metastability' in connection with the von Neumann Mean Ergodic Theorem (MET).
- 2007, independently from Tao, the first rate of asymptotic regularity for MET was extracted in Avigad, Gerhardy, and Towsner [2010].

- 2008, Kohlenbach and Leuştean [2009] generalized this with a better bound to uniformly convex Banach spaces which, subsequently, led to oscillation bounds by Avigad and Rute [2015] (see below).

- Since then, many papers extracting explicit rates of metastability have appeared, including Avigad and Rute [2015], Kohlenbach [2011, 2012, 2016], Kohlenbach and Koutsoukou-Argyraki [2015], Kohlenbach and Leuştean [2009, 2012, 2014], Kohlenbach, Leuştean, and Nicolae [2018], Kohlenbach, López-Acedo, and Nicolae [2017b], Körnlein [2016], Leuştean and Nicolae [2016], Safarik [2012], and Sipoş [2017a].

We like to emphasize that sometimes in analyzing convergence proofs one uses a combination of the approach used in the semi-constructive context discussed further above (applied to those parts of the proof that do not require Σ^0_1-LEM) and the approach to proofs based on full classical logic (applied to the more noneffective parts of the proof). E.g. Leuştean [2014] and Sipoş [2017b] provide interesting instances of such a hybrid approach.

In very special, but important, cases for applications one can extract rates of convergence for iterative procedures towards some non-unique zero of F or fixed point of T, namely when one has an effective so-called modulus of regularity which is closely related to the concepts of weak sharp minima and metric regularity used in convex optimization (see Kohlenbach, López-Acedo, and Nicolae [2017a]).

3.3 Inclusions between sets of solutions. Consider functions $F, G : X \to \mathbb{R}$ on a metric space (X, d) such that every zero of F is also one of G :

$$\forall x \in X\ (F(x) = 0 \to G(x) = 0)$$

which can be re-written in $\forall\exists$-form as

$$\forall x \in X\ \forall k \in \mathbb{N}\ \exists n \in \mathbb{N}\ \overbrace{(|F(x)| \le 2^{-n} \to |G(x)| < 2^{-k})}^{\in\Sigma^0_1}$$

so that logical metatheorems can be applied to extract effective uniform bounds (which due to monotonicity are in fact realizers) for '$\exists n$', i.e.

$$\forall k \in \mathbb{N}\ (|F(x)| \le 2^{-\Phi(x^*,k)} \to |G(x)| < 2^{-k}),$$

where x^* are appropriate majorizing data for x.
For concrete instances of such applications see sections 4 and 6 below.

3.4 Extraction of effective moduli. The first applications of the proof mining methodology in analysis concerned the extraction of explicit moduli of uniqueness in the aforementioned sense (as well as so-called constants of strong unicity) in Chebycheff approximation by us in 1990-1993 which in 2003 - together with Paulo Oliva - was also carried out for best L^1-approximation (see Kohlenbach [2008a] for an extensive coverage of this and the references given there). However, many more concepts of quantitative 'moduli' exist in mathematics or have been introduced as quantitative proof-theoretic versions of qualitative concepts in analysis. Proof mining has been used to explicitly transform moduli for one situation into moduli for another one. This e.g. is used essentially in Bačák and Kohlenbach [2018] and Kohlenbach, López-Acedo, and Nicolae [2017a].

In the rest of the paper we give a few typical examples of explicit bounds which have been obtained by the proof-theoretic machinery discussed so far. For more comprehensive surveys, see Kohlenbach [2008b] for results up to 2008 and Kohlenbach [2017] for applications since 2008.

4 Proof Mining in Convex Analysis

A polynomial rate of asymptotic regularity in Bauschke's solution of the 'zero displacement conjecture'

Consider a real Hilbert space H and nonempty closed and convex subsets $C_1, \ldots, C_N \subseteq H$ with metric projections P_{C_i}, define $T := P_{C_N} \circ \ldots \circ P_{C_1}$. In 2003, Bauschke proved the 'zero displacement conjecture' (Bauschke [2003]) which was first stated in Bauschke, Borwein, and Lewis [1997]:

$$\|T^{n+1}x - T^n x\| \to 0 \quad (x \in H).$$

Previously, this was only known for $N = 2$ or $Fix(T) \neq \emptyset$ (or even $\bigcap_{i=1}^N C_i \neq \emptyset$) or C_i half spaces etc.

The proof uses the Bruck and Reich [1977] theory of firmly and strongly nonexpansive mappings and the abstract theory of maximal monotone operators: Minty's theorem, Brézis-Haraux theorem, Rockafellar's maximal monotonicity and sum theorems, conjugate functions, normal cone operator.

The sequence $(\|T^{n+1}x - T^n x\|)_{n\in\mathbb{N}}$ is nonincreasing and hence the conclusion in Bauschke's theorem is of the form $\forall\exists$. Logical metatheorems as discussed above, therefore, guarantee (modulo the formalizability of the proof in the resp. formal system which, however, does not need to be checked if one explicitly has carried out the extraction) the extractability of an effective uniform rate of asymptotic regularity which only depends on the error $\varepsilon > 0$, $N \in \mathbb{N}$ and majorants for $x \in H$ and $P_{C_1}, \ldots, P_{C_N}$, i.e. $b \geq \|x\|$ and

$K \geq \|c_1\|, \ldots, \|c_N\|$ for some points $c_1 \in C_1, \ldots, c_N \in C_N$ since

$$n \geq \|y\| \to n + K \geq \|P_{C_i} y - P_{C_i} 0\| + \|P_{C_i} 0\| \geq \|P_{C_i} y\|.$$

So one gets a computable $\Phi(\varepsilon, N, b, K)$ s.t. for $b \geq \|x\|$

$$\forall \varepsilon > 0 \, \forall n \geq \Phi(\varepsilon, N, b, K) \, (\|T^{n+1}x - T^n x\| < \varepsilon).$$

Strongly nonexpansive mappings

Definition 4.1 (Kohlenbach [2016]). *Let $S \subseteq X$ be a nonempty subset of a normed space X. $T : S \to X$ is strongly nonexpansive with SNE-modulus $\omega : \mathbb{R}^*_+ \times \mathbb{R}^*_+ \to \mathbb{R}^*_+$ if*

$$\forall d, \varepsilon > 0 \, \forall x, y \in S \, (\|x - y\| \leq d \ \wedge \|x - y\| - \|Tx - Ty\| < \omega(d, \varepsilon) \to \|(x - y) - (Tx - Ty)\| < \varepsilon).$$

Remark: T is strongly nonexpansive in the sense of Bruck and Reich [1977] iff it possesses an SNE-modulus.

Recall that in Hilbert spaces $H = X$, a function $T : S \to H$ is called firmly nonexpansive if

$$\forall x, y \in S \, (\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle)$$

and metric projections onto closed convex subsets of H are firmly nonexpansive.

The next two results have been obtained by a proof-theoretic analysis of Bruck and Reich [ibid.]:

Lemma 4.2 (Kohlenbach [2016]). *Let H be a real Hilbert space and $T = T_N \circ \ldots \circ T_1$ with firmly nonexpansive $T_1, \ldots, T_N : H \to H$. Then T is SNE with modulus*

$$\omega_T(d, \varepsilon) := \frac{1}{16d} \left(\frac{\varepsilon}{N}\right)^2.$$

A rate of asymptotic regularity for SNE-mappings

Theorem 4.3 (Kohlenbach [2018]). *Let $T : S \to S$ be SNE with modulus ω s.t.* $\inf\{\|x - Tx\| : x \in S\} = 0$ *and let $\alpha : \mathbb{R}^*_+ \to \mathbb{R}^*_+$ be such that*

$$\forall \varepsilon > 0 \, \exists y \in S \ \ (\|y\| \leq \alpha(\varepsilon) \wedge \|y - Ty\| \leq \varepsilon) \, .$$

Then for $x \in S$, $x_n := T^n x$ and $D > 0$ such that $\|x - Tx\| \leq D$ one has

$$\forall \varepsilon > 0 \, \forall n \geq \psi(\varepsilon, b, D, \alpha, \omega) \ \ (\|x_{n+1} - x_n\| < \varepsilon) \, , \text{ where}$$

$$\psi(\varepsilon, b, D, \alpha, \omega) := \left\lceil \frac{18b + 12\alpha(\varepsilon/6)}{\varepsilon} - 1 \right\rceil \left\lceil \left(\frac{D}{\omega(D, \tilde{\varepsilon})}\right) \right\rceil, \ \tilde{\varepsilon} := \frac{\varepsilon^2}{27b + 18\alpha(\varepsilon/6)}.$$

The proof-theoretic analysis of the **operator-theoretic part of Bauschke's proof** gives:

Theorem 4.4 (Kohlenbach [2018]). *Let H be real Hilbert space, $C_1, \ldots, C_N \subseteq H$ nonempty closed and convex subsets, P_{C_i} metric projections onto C_i for $i = 1, \ldots, N$. Let $c = (c_1, \ldots, c_N) \in C_1 \times \ldots \times C_N$ be arbitrary and $K \geq \|c\| = \sqrt{\sum_{i=1}^N \|c_i\|^2}$. Let $T := P_{C_N} \circ \ldots \circ P_{C_1}$. Then for every $\varepsilon \in (0,1)$ there exists a point $y \in C_N$ with*

$$\|y\| \leq \alpha(\varepsilon) \text{ and } \|Ty - y\| \leq \varepsilon, \text{ where}$$

$$\alpha(\varepsilon) := \frac{(K^2 + N^3(N-1)^2K^2)N^2}{\varepsilon}.$$

Corollary 4.5 (Kohlenbach [ibid.]).

$$\Phi(\varepsilon, N, b, K) := \left\lceil \frac{18b + 12\alpha(\varepsilon/6))}{\varepsilon} - 1 \right\rceil \left\lceil \left(\frac{D}{\omega(D, \tilde{\varepsilon})} \right) \right\rceil$$

is a rate of asymptotic regularity in Bauschke's result, where

$$\tilde{\varepsilon} := \frac{\varepsilon^2}{27b + 18\alpha(\varepsilon/6)}, \quad D := 2b + NK, \quad \omega(D, \tilde{\varepsilon}) := \frac{1}{16D}(\tilde{\varepsilon}/N)^2.$$

$$\alpha(\varepsilon) := \frac{(K^2 + N^3(N-1)^2K^2)N^2}{\varepsilon}.$$

The case where $Fix(T) \neq \emptyset$ is much simpler:

Theorem 4.6 (Kohlenbach [2016]). *Let $C \subseteq H$ be any nonempty subset of a real Hilbert space H, $T_1, \ldots, T_N : C \to C$ be firmly nonexpansive. Let $T := T_N \circ \ldots \circ T_1$ possess a fixed point $p \in C$ and, for $x \in C$, let $b \geq \|x - p\|, b > 0$. Then for $x_n := T^n x$:*

$$\forall \varepsilon > 0\, \forall n \geq \lceil b/\omega_T(b, \varepsilon) \rceil \, (\|x_{n+1} - x_n\| < \varepsilon), \text{ where}$$

$$\omega_T(b, \varepsilon) := \frac{1}{16b}(\varepsilon/N)^2.$$

Convex feasibility problems

If in Theorem 4.6 the fixed point sets $Fix(T_1), \ldots, Fix(T_N)$ have a nonempty intersection, then any fixed point of T in fact is a common fixed point of $T_1, \ldots, T_N$. This even holds for arbitrary strongly nonexpansive mappings $T_1, \ldots, T_N$ in arbitrary Banach spaces X. In Kohlenbach [ibid.], an explicit bound $\rho(b, \varepsilon)$ (in terms of SNE-moduli for $T_1, \ldots, T_N$) is extracted from the classical proof of this fact such that for $x, p \in C$, p a common fixed point of $T_1, \ldots, T_N$ and $b \geq \|x - p\|$

$$\forall \varepsilon > 0\, (\|T_N T_{N-1} \ldots T_1 x - x\| < \rho(b, \varepsilon) \to \bigwedge_{i=1}^{N} (\|T_i x - x\| < \varepsilon)).$$

Combined with a rate of asymptotic regularity for $T = T_N \circ \ldots \circ T_1$ (which even in this generality is provided in Kohlenbach [ibid.]) this quantitatively solves the problem of constructing a common approximate fixed point of $T_1, \ldots, T_N$.

All this largely holds even in general metric spaces and for strongly quasi-nonexpansive mappings in the sense of Bruck [1982]. Metric projections in so-called CAT(κ)-spaces X (in the sense of Gromov) with $\kappa > 0$ are strongly quasi-nonexpansive and one can construct an explicit modulus for this property which then makes it possible to quantitatively solve the problem to construct a point in the intersection of (ε-neighbourhoods of) finitely many overlapping closed convex subsets of X (i.e. the so-called convex feasibility problem for CAT(κ)-spaces). In the case where X is compact one obtains a rate of metastability for the strong convergence of the iterative use of the composition of the corresponding projections towards a point in the intersection of these sets (see Kohlenbach [2016]).

Other quantitative results in convex optimization have been obtained in

- Ariza-Ruiz, López-Acedo, and Nicolae [2015] and Kohlenbach, López-Acedo, and Nicolae [2017b]: rates of asymptotic regularity and - for compact X - metastability for iterations of compositions of two resolvents in CAT(0)-spaces.
- Kohlenbach, Leuştean, and Nicolae [2018], Kohlenbach, López-Acedo, and Nicolae [2017a], and Sipoş [2017a] rates of asymptotic regularity, strong convergence (in special cases) resp. metastability for the proximal point algorithm.
- Körnlein [2016] explicit such rates for Yamada's hybrid steepest descent method.

5 Proof Mining in Ergodic Theory

Let H be a real Hilbert space, $T : H \to H$ be linear and $\|T(x)\| \leq \|x\|$ for all $x \in H$. Consider the Cesàro mean of the iterates of T :

$$A_n(x) := \frac{1}{n} S_n(x), \text{ where } S_n(x) := \sum_{i=0}^{n-1} T^i(x) \quad (n \geq 1).$$

The von Neumann Mean Ergodic Theorem in the formulation of Riesz states:

Theorem 5.1 (von Neumann Mean Ergodic Theorem)**.** *For every $x \in H$, the sequence $(A_n(x))_n$ strongly converges.*

In Avigad, Gerhardy, and Towsner [2010], it is shown that in general there is no computable rate of convergence, but a primitive recursive rate of metastability is extracted using the proof-theoretic methods discussed above. Tao [2008a] also established (without bound) a uniform metastable version of the Mean Ergodic Theorem in Hilbert space and used that uniformity as a base step for a generalization to commuting families of operators. On the connection to the proof-theoretic approach he comments:

'We shall establish Theorem 1.6 by "finitary ergodic theory" techniques, reminiscent of those used in [Green-Tao]...' 'The main advantage of working in the finitary setting ... is that the underlying dynamical system becomes extremely explicit'...'In proof theory, this finitisation is known as Gödel functional interpretation...which is also closely related to the Kreisel no-counterexample interpretation' (T. Tao [2008a]).
In 1939, Garrett Birkhoff proved:

Theorem 5.2 (Birkhoff [1939]). *The Mean Ergodic Theorem holds for arbitrary uniformly convex Banach spaces.*

Remark 5.3. *In the same year as Birkhoff [ibid.], Lorch [1939] showed that the mean ergodic theorem even holds in all reflexive spaces. However, the class of reflexive spaces does not have enough uniformity to allow for a logical metatheorem on uniform bound extractions and, in fact, in Avigad and Rute [2015] it is shown that a uniform rate of metastability has to depend on the modulus of uniform convexity.*

Since Birkhoff's proof formalizes in the deductive framework of uniformly convex normed spaces (with modulus η) $\mathfrak{A}^\omega[X, \|\cdot\|, \eta]$ (see Kohlenbach [2008a] for the definition of this system) the following is guaranteed a-priorily:

Let X be a uniformly convex Banach space with modulus η and $T : X \to X$ nonexpansive linear operator. Let $b > 0$. Then there is an effective functional Φ in ε, g, b, η s.t. for all $x \in X$ with $\|x\| \le b$, all $\varepsilon > 0$, all $g : \mathbb{N} \to \mathbb{N}$:

$$\exists n \le \Phi(\varepsilon, g, b, \eta)\, \forall i, j \in [n, n + g(n)]\, \big(\|A_i(x) - A_j(x)\| < \varepsilon\big).$$

Note that $T^* := id$ majorizes T.

Based on the logical metatheorem above (for uniformly convex normed spaces) the following rate of metastability was extracted from Birkhoff's proof:

Theorem 5.4 (Kohlenbach and Leuştean [2009]). *Let X be a uniformly convex Banach space, η a modulus of uniform convexity, $T : X \to X$ be as above and $b > 0$. Then for all $x \in X$ with $\|x\| \le b$, all $\varepsilon > 0$ and all $g : \mathbb{N} \to \mathbb{N}$:*

$$\exists n \le \Phi(\varepsilon, g, b, \eta)\, \forall i, j \in [n, n + g(n)]\, \big(\|A_i(x) - A_j(x)\| < \varepsilon\big), \text{ where}$$

$$\Phi(\varepsilon, g, b, \eta) := M \cdot \tilde{h}^{(K)}(1), \text{ with } M := \left\lceil \frac{16b}{\varepsilon} \right\rceil, \gamma := \frac{\varepsilon}{16}\eta\left(\frac{\varepsilon}{8b}\right), \quad K := \left\lceil \frac{b}{\gamma} \right\rceil,$$
$$h, \tilde{h} : \mathbb{N} \to \mathbb{N},\ h(n) := 2(Mn + g(Mn)), \quad \tilde{h}(n) := \max_{i \le n} h(i).$$

If $\eta(\varepsilon) = \varepsilon \cdot \tilde{\eta}(\varepsilon)$ with increasing $\tilde{\eta}$, then we can replace 'η' by '$\tilde{\eta}$' and '16' by '8'. In particular, for $X = L^p$ with $1 < p < \infty$, we may take $\tilde{\eta}(\varepsilon) = \varepsilon^{p-1}/(p2^p)$.

Bounding the number of fluctuations: We say that (x_n) admits k ε-fluctuations if there are $i_1 \leq j_1 \leq \ldots i_k \leq j_k$ s.t. $\|x_{j_n} - x_{i_n}\| \geq \varepsilon$ for $n = 1, \ldots, k$.

Using the analysis of Birkhoff's proof in Kohlenbach and Leuştean [ibid.], Avigad and Rute subsequently improved the rate of metastability to a bound on the number of ε-fluctuations:

Theorem 5.5 (Avigad and Rute [2015]). *$(A_n(x))$ admits at most*

$$\left\lfloor 4\log(M) \cdot \frac{b}{\varepsilon} \right\rfloor + \left\lfloor \frac{b}{\gamma} \right\rfloor \cdot \left\lfloor (4\log(2M) \cdot \frac{b}{\varepsilon} \right\rfloor + \left\lfloor \frac{b}{\gamma} \right\rfloor$$

many ε-fluctuations with b, M, γ as in Theorem 5.4.

In the Hilbert space case, fluctuation bounds had already been obtained in Jones, Ostrovskii, and Rosenblatt [1996].
If the linearity of the nonexpansive operator T is dropped, then the convergence of (x_n) holds weakly (but in general not strongly, see Genel and Lindenstrauss [1975]) by Baillon's nonlinear ergodic theorem:

Theorem 5.6 (Baillon [1975]). *Let H be a real Hilbert space, $C \subseteq H$ bounded closed and convex and $T : C \to C$ be nonexpansive. Then for every $x_0 \in C$, the sequence of Cesàro means (x_n) converges weakly to a fixed point of T.*

A rate of metastability for the weak Cauchy property is extracted in Kohlenbach [2012].

If one either changes the Cesàro means slightly (or adds some weak form of linearity, see below) one can achieve strong convergence. Consider the so-called Halpern iteration Halpern [1967]: Let $T : C \to C$ be nonexpansive, $x_1 \in C$, $\alpha_n \in [0, 1]$

$$x_{n+1} := \alpha_n x_1 + (1 - \alpha_n) T(x_n) \quad (n \geq 1).$$

In contrast to other iterative schemes such as Krasnoselski-Mann iterations, the Halpern iteration often converges strongly (one reason, though, why it is less used convex optimization is that it is not Fejér monotone; see Kohlenbach, Leuştean, and Nicolae [2018] for explicit rates of metastability from strong convergence proofs based on Fejér monotonicity).
Using a weak compactness argument, Wittmann proved in 1992 the following strong convergence result:

Theorem 5.7 (Wittmann [1992]). *Let H be a real Hilbert space, $C \subseteq H$ closed and convex, $x_0 \in C$ and $Fix(T) \neq \emptyset$. Under suitable conditions on (α_n) (e.g. for $\alpha_n := \frac{1}{n+1}$) (x_n) converges strongly towards the fixed point of T that is closest to x_0.*

Remark 5.8. 1. *Wittmann's theorem is a nonlinear generalization of the Mean Ergodic Theorem: for* $\alpha_n := 1/(n+1), C := H$ *and linear* T*, the Halpern iteration coincides with the Cesàro means.*

2. *Another nonlinear generalization of the Mean Ergodic Theorem has been obtained in* Baillon [1976]. *Here one keeps the original Cesàro means but requires that* T *(in addition to being nonexpansive) is odd (and* C *is symmetric). This was further generalized in* Wittmann [1990] *from which an explicit rate of metastability was extracted in* Safarik [2012].

Wittmann's result has been generalized to CAT(0)-spaces by Saejung [2010] using Banach limits. Explicit rates of metastability have been extracted in Kohlenbach [2011] (for Hilbert spaces) with an elimination of the use of weak compactness and in Kohlenbach and Leuştean [2012, 2014] (for CAT(0) spaces) with an elimination of the use of Banach limits.

Moreover, one has a quadratic rate of asymptotic regularity $d(x_n, T(x_n)) \to 0$:

$$\forall \varepsilon > 0\, \forall n \geq \frac{4M}{\varepsilon} + \frac{32M^2}{\varepsilon^2}\ (d(x_n, T(x_n)) < \varepsilon)$$

(See Kohlenbach and Leuştean [2012].) In Leuştean and Nicolae [2016], the proof-theoretic analysis of Saejung's proof has been further generalized to the highly nontrivial case of CAT(κ)-spaces for $\kappa > 0$ producing an explicit rate of metastability even in this context.

6 Proof Mining in Nonlinear Semigroup Theory

Let X be a Banach space, $C \subseteq X$ be a nonempty subset and $\lambda \in (0,1)$.

Definition: A family $\{T(t) : t \geq 0\}$ of nonexpansive mappings $T(t) : C \to C$ is a nonexpansive semigroup if

(i) $T(s+t) = T(s) \circ T(t)\ \ (s,t \geq 0)$,
(ii) for each $x \in C$, the mapping $t \mapsto T(t)x$ is continuous.

Theorem 6.1 (Suzuki [2006]). *Let* $0 < \alpha < \beta$ *such that* α/β *is irrational. Then any fixed point* $p \in C$ *of*

$$S := \lambda T(\alpha) + (1-\lambda)T(\beta) : C \to X$$

is a common fixed point of $T(t)$ *for all* $t \geq 0$.

Let $t \mapsto T(t)x$ be equicontinuous on norm-bounded subsets of C with modulus ω, let f_γ be an effective irrationality measure for $\gamma := \alpha/\beta$, $\Lambda, N, D \in \mathbb{N}$ be s.t. $1/\Lambda \leq \lambda, 1-\lambda$

and $1/N \le \beta \le D$. Then one can extract a bound (see Section 3.3) $\Phi(\varepsilon, M, b) := \Phi(\varepsilon, M, b, N, \Lambda, D, f_\gamma, \omega)$ s.t. for all $M, b \in \mathbb{N}$, $p \in C$, $\varepsilon > 0$

$$\|p\| \le b \wedge \|S(p) - p\| \le \Phi(\varepsilon, M, b) \to \forall t \in [0, M]\,(\|T(t)p - p\| \le \varepsilon).$$

The main noneffective tool used in Suzuki's proof is the binary König's lemma WKL and by Remark 2.4 it is guaranteed to have a primitive recursive (in the sense of Kleene) bound Φ. In fact, the bound actually extracted in Kohlenbach and Koutsoukou-Argyraki [2016] is of rather low complexity:

$$\Phi(2^{-m}, M, b) = \frac{2^{-m}}{8(\sum_{i=1}^{\phi(k,f_\gamma)-1} \Lambda^i + 1)(1 + MN)}, \text{ where}$$

$$k := D2^{\omega_{D,b}(3+[\log_2(1+MN)]+m)+1},\ \phi(k, f) := \max\{2f(i) + 6 \,:\, 0 < i \le k\}.$$

Example: $\alpha = \sqrt{2}, \beta = 2, \lambda = 1/2$. Then $\Lambda = 2, N = 1, D = 2, f_\gamma(p) = 4p^2$. If C is convex (so that $S : C \to C$) and $x_{n+1} := \frac{1}{2}x_n + \frac{1}{2}Sx_n \in C$ starting from $x_0 \in C$ is a d-bounded Krasnoselski iteration sequence of S one has a quadratic rate of asymptotic regularity $\Psi(\varepsilon, d) := 4d^2/(\pi\varepsilon^2)$ (Baillon and Bruck [1996]) and so

$$\forall n \ge \Psi(\Phi(\varepsilon, M, b), d)\ \forall t \in [0, M]\,(\|T(t)x_n - x_n\| \le \varepsilon).$$

Nonexpansive semigroups feature prominently - via the Crandall-Liggett formula - in the study of abstract Cauchy problems that are given by accretive set-valued operators. Explicit rates on the asymptotic behavior of solutions have been obtained by our proof-theoretic methods in Kohlenbach and Koutsoukou-Argyraki [2015] and Koutsoukou-Argyraki [2017].

References

David Ariza-Ruiz, Genaro López-Acedo, and Adriana Nicolae (2015). "The asymptotic behavior of the composition of firmly nonexpansive mappings". *J. Optim. Theory Appl.* 167.2, pp. 409–429. MR: 3412444 (cit. on p. 93).

Jeremy Avigad, Philipp Gerhardy, and Henry Towsner (2010). "Local stability of ergodic averages". *Trans. Amer. Math. Soc.* 362.1, pp. 261–288. MR: 2550151 (cit. on pp. 88, 93).

Jeremy Avigad and Jason Rute (2015). "Oscillation and the mean ergodic theorem for uniformly convex Banach spaces". *Ergodic Theory Dynam. Systems* 35.4, pp. 1009–1027. MR: 3345161 (cit. on pp. 88, 89, 94, 95).

Miroslav Bačák and Ulrich Kohlenbach (2018). "On proximal mappings with Young functions in uniformly convex Banach spaces". *J. Convex Analysis* 25 (4) (cit. on p. 90).

Jean-Bernard Baillon (1975). "Un théorème de type ergodique pour les contractions non linéaires dans un espace de Hilbert". *C. R. Acad. Sci. Paris Sér. A-B* 280.22, Aii, A1511–A1514. MR: 0375009 (cit. on p. 95).

– (1976). "Quelques propriétés de convergence asymptotique pour les contractions impaires". *C. R. Acad. Sci. Paris Sér. A-B* 283.8, Aii, A587–A590. MR: 0428124 (cit. on p. 96).

Jean-Bernard Baillon and R. E. Bruck (1996). "The rate of asymptotic regularity is $O(1/n)$ $[O(1/\surd$ In: *Theory and applications of nonlinear operators of accretive and monotone type*. Vol. 178. Lecture Notes in Pure and Appl. Math. Dekker, New York, pp. 51–81. MR: 1386668 (cit. on p. 97).

Heinz H. Bauschke (2003). "The composition of projections onto closed convex sets in Hilbert space is asymptotically regular". *Proc. Amer. Math. Soc.* 131.1, pp. 141–146. MR: 1929033 (cit. on p. 90).

Heinz H. Bauschke, Jonathan M. Borwein, and Adrian S. Lewis (1997). "The method of cyclic projections for closed convex sets in Hilbert space". In: *Recent developments in optimization theory and nonlinear analysis (Jerusalem, 1995)*. Vol. 204. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 1–38. MR: 1442992 (cit. on p. 90).

Garrett Birkhoff (1939). "The mean ergodic theorem". *Duke Math. J.* 5.1, pp. 19–20. MR: 1546101 (cit. on p. 94).

E. Bombieri and A. J. van der Poorten (1988). "Some quantitative results related to Roth's theorem". *J. Austral. Math. Soc. Ser. A* 45.2, pp. 233–248. MR: 951583 (cit. on p. 81).

Eyvind Martol Briseid (2009). "Logical aspects of rates of convergence in metric spaces". *J. Symbolic Logic* 74.4, pp. 1401–1428. MR: 2583826 (cit. on p. 87).

F. E. Browder and W. V. Petryshyn (1966). "The solution by iteration of nonlinear functional equations in Banach spaces". *Bull. Amer. Math. Soc.* 72, pp. 571–575. MR: 0190745 (cit. on p. 85).

Ronald E. Bruck (1982). "Random products of contractions in metric and Banach spaces". *J. Math. Anal. Appl.* 88.2, pp. 319–332. MR: 667060 (cit. on p. 93).

Ronald E. Bruck and Simeon Reich (1977). "Nonexpansive projections and resolvents of accretive operators in Banach spaces". *Houston J. Math.* 3.4, pp. 459–470. MR: 0470761 (cit. on pp. 90, 91).

Simon Cho (2016). "A variant of continuous logic and applications to fixed point theory". arXiv: 1610.05397 (cit. on p. 82).

A. Genel and J. Lindenstrauss (1975). "An example concerning fixed points". *Israel J. Math.* 22.1, pp. 81–86. MR: 0390847 (cit. on p. 95).

Gerhard Gentzen (1936). "Die Widerspruchsfreiheit der reinen Zahlentheorie". *Math. Ann.* 112.1, pp. 493–565. MR: 1513060 (cit. on p. 80).

Philipp Gerhardy and Ulrich Kohlenbach (2008). "General logical metatheorems for functional analysis". *Trans. Amer. Math. Soc.* 360.5, pp. 2615–2660. MR: 2373327 (cit. on pp. 83, 84).

Kurt Gödel (1941). *In what sense is intuitionistic logic constructive*. Lecture at Yale. First published on pages 189–200 of Gödel [1995] (cit. on pp. 80, 81).

– (1958). "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes". *Dialectica* 12, pp. 280–287. MR: 0102482 (cit. on pp. 80, 81).

– (1995). *Collected works. Vol. III – Unpublished essays and lectures*. Edited and with a preface by Solomon Feferman. The Clarendon Press, Oxford University Press, New York, pp. xvi+474. MR: 1332489.

Kazimierz Goebel and W. A. Kirk (1983). "Iteration processes for nonexpansive mappings". In: *Topological methods in nonlinear functional analysis (Toronto, Ont., 1982)*. Vol. 21. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 115–123. MR: 729507 (cit. on p. 83).

Daniel Günzel and Ulrich Kohlenbach (2016). "Logical metatheorems for abstract spaces axiomatized in positive bounded logic". *Adv. Math.* 290, pp. 503–551. MR: 3451931 (cit. on p. 81).

Benjamin Halpern (1967). "Fixed points of nonexpanding maps". *Bull. Amer. Math. Soc.* 73, pp. 957–961. MR: 0218938 (cit. on p. 95).

David Hilbert (1926). "Über das Unendliche". *Math. Ann.* 95.1, pp. 161–190. MR: 1512272 (cit. on p. 80).

W. A. Howard (1973). "Appendix: Hereditarily majorizable functionals of finite type". In: *Metamathematical investigation of intuitionistic arithmetic and analysis*. Ed. by A. S. Troelstra. Vol. 344. Lecture Notes in Math. Springer, Berlin, pp. 454–461. MR: 0469712 (cit. on p. 83).

Roger L. Jones, Iosif V. Ostrovskii, and Joseph M. Rosenblatt (1996). "Square functions in ergodic theory". *Ergodic Theory Dynam. Systems* 16.2, pp. 267–305. MR: 1389625 (cit. on p. 95).

W. A. Kirk (1981/82). "Krasnoselskiĭ's iteration process in hyperbolic space". *Numer. Funct. Anal. Optim.* 4.4, pp. 371–381. MR: 673318 (cit. on p. 83).

Ulrich Kohlenbach (1993). "Effective moduli from ineffective uniqueness proofs. An unwinding of de la Vallée Poussin's proof for Chebycheff approximation". *Ann. Pure Appl. Logic* 64.1, pp. 27–94. MR: 1241250 (cit. on p. 86).

– (2005a). "Some computational aspects of metric fixed-point theory". *Nonlinear Anal.* 61.5, pp. 823–837. MR: 2130066 (cit. on p. 88).

– (2005b). "Some logical metatheorems with applications in functional analysis". *Trans. Amer. Math. Soc.* 357.1, pp. 89–128. MR: 2098088 (cit. on pp. 81, 83, 84).

Ulrich Kohlenbach (2008a). *Applied proof theory: proof interpretations and their use in mathematics*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, pp. xx+532. MR: 2445721 (cit. on pp. 81–84, 86, 88, 90, 94).

– (2008b). "Effective uniform bounds from proofs in abstract functional analysis". In: *New computational paradigms*. Springer, New York, pp. 223–258. MR: 2762087 (cit. on pp. 85, 90).

– (2011). "On quantitative versions of theorems due to F. E. Browder and R. Wittmann". *Adv. Math.* 226.3, pp. 2764–2795. MR: 2739793 (cit. on pp. 89, 96).

– (2012). "A uniform quantitative form of sequential weak compactness and Baillon's nonlinear ergodic theorem". *Commun. Contemp. Math.* 14.1, 20 pp. MR: 2902296 (cit. on pp. 89, 95).

– (2016). "On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces". *Israel J. Math.* 216.1, pp. 215–246. MR: 3556967 (cit. on pp. 88, 89, 91–93).

– (2017). "Recent progress in proof mining in nonlinear analysis". *IFCoLog Journal of Logics and their Applications* 10 (4), pp. 3357–3406 (cit. on pp. 85, 90).

– (2018). "A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space". *Found Comput Math* (cit. on pp. 91, 92).

Ulrich Kohlenbach and Angeliki Koutsoukou-Argyraki (2015). "Rates of convergence and metastability for abstract Cauchy problems generated by accretive operators". *J. Math. Anal. Appl.* 423.2, pp. 1089–1112. MR: 3278188 (cit. on pp. 89, 97).

– (2016). "Effective asymptotic regularity for one-parameter nonexpansive semigroups". *J. Math. Anal. Appl.* 433.2, pp. 1883–1903. MR: 3398796 (cit. on p. 97).

Ulrich Kohlenbach and Branimir Lambov (2004). "Bounds on iterations of asymptotically quasi-nonexpansive mappings". In: *International Conference on Fixed Point Theory and Applications, Valencia 2003*. Yokohama Publ., Yokohama, pp. 143–172. MR: 2144170 (cit. on p. 88).

Ulrich Kohlenbach and Laurenţiu Leuştean (2009). "A quantitative mean ergodic theorem for uniformly convex Banach spaces". *Ergodic Theory Dynam. Systems* 29.6, pp. 1907–1915. MR: 2563097 (cit. on pp. 89, 94, 95).

– (2012). "Effective metastability of Halpern iterates in CAT(0) spaces". *Adv. Math.* 231, pp. 2526–2556. MR: 2970458 (cit. on pp. 89, 96).

– (2014). "Addendum to "Effective metastability of Halpern iterates in CAT(0) spaces" [Adv. Math. 231 (5) (2012) 2526–2556] [MR2970458]". *Adv. Math.* 250, pp. 650–651. MR: 3122180 (cit. on pp. 89, 96).

Ulrich Kohlenbach, Laurenţiu Leuştean, and Adriana Nicolae (2018). "Quantitative results on Fejér monotone sequences". Comm. Contemp. Math 20, 42 pp. (cit. on pp. 88, 89, 93, 95).

Ulrich Kohlenbach, Genaro López-Acedo, and Adriana Nicolae (2017a). "Moduli of regularity and rates of convergence for Fejér monotone sequences". Preprint. arXiv: 1711.02130 (cit. on pp. 89, 90, 93).

– (2017b). "Quantitative asymptotic regularity results for the composition of two mappings". *Optimization* 66.8, pp. 1291–1299. MR: 3665825 (cit. on pp. 89, 93).

Ulrich Kohlenbach and Pavol Safarik (2014). "Fluctuations, effective learnability and metastability in analysis". *Ann. Pure Appl. Logic* 165.1, pp. 266–304. MR: 3111912 (cit. on p. 88).

Daniel Körnlein (2016). "Quantitative Analysis of Iterative Algorithms in Fixed Point Theory and Convex Optimization". PhD thesis. Technische Universität Darmstadt (cit. on pp. 89, 93).

Angeliki Koutsoukou-Argyraki (2017). "Effective rates of convergence for the resolvents of accretive operators". *Numer. Funct. Anal. Optim.* 38.12, pp. 1601–1613. MR: 3740740 (cit. on p. 97).

Georg Kreisel (1951). "On the interpretation of non-finitist proofs. I". *J. Symbolic Logic* 16, pp. 241–267. MR: 0049135 (cit. on pp. 80, 87).

– (1952). "On the interpretation of non-finitist proofs. II. Interpretation of number theory. Applications". *J. Symbolic Logic* 17, pp. 43–58. MR: 0051193 (cit. on pp. 80, 87).

– (1959). "Interpretation of analysis by means of constructive functionals of finite types". In: *Constructivity in mathematics: Proceedings of the colloquium held at Amsterdam, 1957 (edited by A. Heyting)*. Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Co., Amsterdam, pp. 101–128. MR: 0106838 (cit. on p. 80).

– (1982). "Finiteness theorems in arithmetic: an application of Herbrand's theorem for Σ_2-formulas". In: *Proceedings of the Herbrand symposium (Marseilles, 1981)*. Vol. 107. Stud. Logic Found. Math. North-Holland, Amsterdam, pp. 39–55. MR: 757021 (cit. on p. 81).

Laurenţiu Leuştean (2014). "An application of proof mining to nonlinear iterations". *Ann. Pure Appl. Logic* 165, pp. 1484–1500. MR: 3210080 (cit. on p. 89).

Laurenţiu Leuştean and Adriana Nicolae (2016). "Effective results on nonlinear ergodic averages in CAT(κ) spaces". *Ergodic Theory Dynam. Systems* 36.8, pp. 2580–2601. MR: 3570025 (cit. on pp. 89, 96).

Edgar R. Lorch (1939). "Means of iterated transformations in reflexive vector spaces". *Bull. Amer. Math. Soc.* 45, pp. 945–947. MR: 0001460 (cit. on p. 94).

H. Luckhardt (1989). "Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken". *J. Symbolic Logic* 54.1, pp. 234–263. MR: 987335 (cit. on p. 81).

Eike Neumann (2015). "Computational problems in metric fixed point theory and their Weihrauch degrees". *Log. Methods Comput. Sci.* 11.4, 44 pp. MR: 3438729 (cit. on p. 87).

Charles Parsons (1972). "On n-quantifier induction". *J. Symbolic Logic* 37, pp. 466–482. MR: 0325365 (cit. on p. 80).

Simeon Reich and Itai Shafrir (1990). "Nonexpansive iterations in hyperbolic spaces". *Nonlinear Anal.* 15.6, pp. 537–558. MR: 1072312 (cit. on p. 83).

Satit Saejung (2010). "Halpern's iteration in CAT(0) spaces". *Fixed Point Theory Appl.* 2010. Article ID 471781, 13 pp. MR: 2578767 (cit. on p. 96).

Pavol Safarik (2012). "A quantitative nonlinear strong ergodic theorem for Hilbert spaces". *J. Math. Anal. Appl.* 391.1, pp. 26–37. MR: 2899834 (cit. on pp. 89, 96).

Andrei Sipoş (2017a). "Contributions to Proof Mining". PhD thesis. Univ. of Bucharest (cit. on pp. 89, 93).

– (2017b). "Effective results on a fixed point algorithm for families of nonlinear mappings". *Ann. Pure Appl. Logic* 168.1, pp. 112–128. MR: 3564378 (cit. on p. 89).

Craig Smorynski (1977). "The incompleteness theorems". In: *Barwise, J. (ed.), Handbook of Mathematical Logic*, pp. 821–865 (cit. on p. 79).

Ernst Specker (1949). "Nicht konstruktiv beweisbare Sätze der Analysis". *J. Symbolic Logic* 14, pp. 145–158. MR: 0031447 (cit. on pp. 85, 86).

Tomonari Suzuki (2006). "Common fixed points of one-parameter nonexpansive semigroups". *Bull. London Math. Soc.* 38.6, pp. 1009–1018. MR: 2285255 (cit. on p. 96).

Wataru Takahashi (1970). "A convexity in metric space and nonexpansive mappings. I". *Kōdai Math. Sem. Rep.* 22, pp. 142–149. MR: 0267565 (cit. on p. 83).

Terence Tao (2008a). "Norm convergence of multiple ergodic averages for commuting transformations". *Ergodic Theory Dynam. Systems* 28.2, pp. 657–688. MR: 2408398 (cit. on pp. 93, 94).

– (2008b). *Structure and randomness*. Pages from year one of a mathematical blog. American Mathematical Society, Providence, RI, pp. xii+298. MR: 2459552 (cit. on pp. 87, 88).

Rainer Wittmann (1990). "Mean ergodic theorems for nonlinear operators". *Proc. Amer. Math. Soc.* 108.3, pp. 781–788. MR: 1004427 (cit. on p. 96).

– (1992). "Approximation of fixed points of nonexpansive mappings". *Arch. Math. (Basel)* 58.5, pp. 486–491. MR: 1156581 (cit. on p. 95).

Received 2017-12-05.

Ulrich Kohlenbach
Department of Mathematics
Technische Universität Darmstadt
Schlossgartenstraße 7
64289 Darmstadt
Germany
kohlenbach@mathematik.tu-darmstadt.de

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (101–116)

MODEL THEORY AND ULTRAPRODUCTS

Maryanthe Malliaris

Abstract

The article motivates recent work on saturation of ultrapowers from a general mathematical point of view.

Introduction

In the history of mathematics the idea of the limit has had a remarkable effect on organizing and making visible a certain kind of structure. Its effects can be seen from calculus to extremal graph theory.

At the end of the nineteenth century, Cantor introduced infinite cardinals, which allow for a stratification of size in a potentially much deeper way. It is often thought that these ideas, however beautiful, are studied in modern mathematics primarily by set theorists, and that aside from occasional independence results, the hierarchy of infinities has not profoundly influenced, say, algebra, geometry, analysis or topology, and perhaps is not suited to do so.

What this conventional wisdom misses is a powerful kind of reflection or transmutation arising from the development of model theory. As the field of model theory has developed over the last century, its methods have allowed for serious interaction with algebraic geometry, number theory, and general topology. One of the unique successes of model theoretic classification theory is precisely in allowing for a kind of distillation and focusing of the information gleaned from the opening up of the hierarchy of infinities into definitions and tools for studying specific mathematical structures.

By the time they are used, the form of the tools may not reflect this influence. However, if we are interested in advancing further, it may be useful to remember it. Along this seam, so to speak, may be precisely where we will want to re-orient our approach.

Partially supported by NSF 1553653, and at IAS by NSF 1128155 and a Minerva Research Foundation membership.
MSC2010: primary 03C45; secondary 03C20, 03E17, 05C99.

1 The model theoretic point of view

The model theoretic setup is designed to allow in a specific way for placing a given infinite mathematical structure in a class or family in which size and certain other features, such as the appearance of limit points, may vary.

Suppose we wanted to look abstractly at the structure the reals carry when viewed as an ordered field. We might consider $\mathbb{R}$ as a set decorated by the following: a directed graph edge representing the ordering, a first directed hyperedge representing the graph of addition, a second directed hyperedge representing the graph of multiplication, perhaps a third for the graph of subtraction, and two constant symbols marking the additive and multiplicative identities. Stepping back, and retaining only this data of a set of size continuum along with the data of which sets of elements or tuples correspond to which constants, edges, or hyperedges, we might try to analyze the configurations which do or do not arise there.

Suppose we were interested in the structure the group operation gives to the discrete Heisenberg group H. We might consider H simply as an infinite set along with the data of the multiplication table. A priori, this setup just records a countably infinite set made into a group in the given way; it doesn't a priori record that its elements are matrices, much less uni-upper-triangular matrices over $\mathbb{Z}$.

These examples suggest how models arise – simply as sets decorated by the data of relations or functions we single out for study.[1]

The initial loss of information in such a representation will be balanced by the fact that it allows us to place a model within a class and to study models in the class alongside each other. From the model theoretic point of view – the following statement is a starting point for investigation, not its conclusion – this class contains all other models which differ from M in inessential ways.

To place our model M in its class, we consider the *theory* of the model, that is, the set of all sentences of first order logic which hold in M. The elementary class of M is the class of all other models with the same theory. We may frame our study as: of such classes, or of theories.

Every model carries what are essentially derived relations, the boolean algebras of definable sets (see the Appendix). We might say very informally that the theory of a model takes a photograph of these boolean algebras which remembers only finitary information,

[1] A k-place relation on a set X is a subset of X^k. The set of first coordinates of a binary relation is called its domain. A k-place function on a set X is a single-valued binary relation whose domain is X^k. To specify a model, we first choose a language, which can include relation symbols, function symbols, and constant symbols. Then a model is given by the following data: a set X, called the domain; for each relation symbol, a relation on X of the right arity; for each function symbol, a function on X of the right arity; for each constant symbol, an element of X.

such as which finite intersections are and are not empty. The models which share the theory of M will also share this photograph.

The differences among models with the same theory come essentially from the infinite intersections which are finitely approximated, in other words, with the filters and ultrafilters on the boolean algebra of definable sets, as we now explain.

2 Limit points

To see what engenders variation within an elementary class, the following story will motivate the definition of *type*. (The motivation is in the telling, not an historical assertion. Filters independently arose in other, earlier contexts in the early part of the twentieth century.)

Writing their book on general topology in 1937, Bourbaki were discussing whether the definition of limit could be liberated from the countable. Cartan's suggestion was effectively that working in a topological space X, one might turn the problem around and look from the point of view of a limit point. Given a point $x \in X$, the family of neighborhoods containing x had certain very nice features – such as upward closure, closure under finite intersection – which may be abstracted as the definition of *filter*.[2]

Definition 2.1. *For I an infinite set, $\mathfrak{F} \subseteq \mathcal{P}(I)$ is a* filter *on I when (i) $A \subseteq B \subseteq I$ and $A \in \mathfrak{F}$ implies $B \in \mathfrak{F}$, (ii) $A, B \in \mathfrak{F}$ implies $A \cap B \in \mathfrak{F}$, (iii) $\emptyset \notin \mathfrak{F}$.*

Conversely, to any filter, one can assign a (possibly empty) set of limit points: those elements of I which belong to all $A \in \mathfrak{F}$. In defining a filter, we may restrict to any boolean algebra $\mathfrak{B} \subseteq \mathcal{P}(I)$, asking that $\mathfrak{F} \subseteq \mathfrak{B}$, and adding that B in item (ii) belong to $\mathfrak{B}$. In the model theoretic context, this idea gives us a natural way to define limit points for any model, not requiring a metric or a topology per se:

Definition 2.2. *Informally, a* partial type p over a model M *is a filter on M for the boolean algebra of M-definable subsets of M. [More correctly, it is a set of formulas with parameters from M, whose solution sets in M form such a filter.*[3]*] It is a* type *if it*

[2] I learned this story from Maurice Mashaal's biography of Bourbaki, which also cites biographical work of Liliane Beaulieu. Regarding Cartan's definition of filters: "At first [the others] met the idea with skepticism, but Chevalley understood the importance of Cartan's suggestion and even proposed another idea based on it (which became the concept of ultrafilters). Once the approval was unanimous, someone yelled « boum ! » (French for "bang!") to announce that a breakthrough had been made – this was one of Bourbaki's many customs." Mashaal [2006]

[3] This difference is visible in the idea of realizing a given type over M in a model N extending M. The new limit point will belong to the solution sets, in N, of the formulas in the type. Note that 2.2 describes types of *elements*, corresponding to sets of formulas in one free variable plus parameters from M. For each $n > 1$ there is an analogous $\mathbf{S}_n(M)$ describing types of n-tuples.

is maximal, i.e. not strictly contained in any other partial type over M. The Stone space $\mathbf{S}(M)$ is the set of types over M.

A type is *realized* if a corresponding limit point exists in the model, otherwise it is *omitted*. For example, if $M = (\mathbb{Q}, <)$ is the rationals considered as a linear order, $\mathbf{S}(M)$ includes a type for each element of M, which are realized, and types for each irrational Dedekind cut, for $+\infty$, for $-\infty$, and for various infinitessimals, which are omitted.

The *compactness theorem* for first order logic ensures that for any model M and any type or set of types over M we can always find an extension of M to a larger model in the same class in which all these types are realized. (Put otherwise, we may realize types without changing the theory.) Types are fundamental objects in all that follows. From the depth and subtlety of their interaction comes much of the special character of the subject.[4]

In Definition 2.2, we may use "A-definable sets" instead of "M-definable sets" for some $A \subseteq M$ [more correctly, formulas in one free variable with parameters from A]. In this case, call p a type or partial type *of M over A*. Then the following fundamental definition, from work of Morley and of Vaught in the early 1960s, generalizing ideas from Hausdorff on η_α sets, gives a measure of the completeness of a model.

Definition 2.3. *For an infinite cardinal κ, we say a model N is κ^+-*saturated *if every type of N over every $A \subseteq N$ of size $\leq \kappa$ is realized in N.*

For example, if M is an algebraically closed field then for any subfield K of M, the types of M over K will include a distinct type for each minimal polynomial over K (describing a root) and one type describing an element transcendental over K. Since it is algebraically closed, M will be κ^+-saturated if and only if it has transcendence degree at least κ^+.

3 Towards classification theory

Having described a model theoretic point of view – first, regarding a given mathematical object as a model; second, placing it within an elementary class of models sharing the same first-order theory; third, studying as our basic objects these theories, looking both at how models may vary for a given theory (by paying close attention to the structure of types) and at structural differences across theories – some first notable features of this setup are:

[4]And a certain possible conversion of combinatorial into algebraic information, as in the remarkable group configuration theorems of Zilber and Hrushovski.

a) *one can study the truth of statements of first order logic by 'moving' statements among models in allowed ways*.[5]

b) *when working in a given elementary class, unusual constraints observed on variation of models may give leverage for a structural understanding of all models in the class*. E.g. Morley's theorem, the 'cornerstone' of modern model theory, says that if a countable theory has only one model up to isomorphism in *some* uncountable size κ, then this must be true in every uncountable size. Moreover its models must behave analogously to algebraically closed fields of a given characteristic in the sense that, e.g., there are prime models over sets, there are relatively few types, and for each model there is a single invariant, a dimension (the equivalent of transcendence degree) giving the isomorphism type.

c) *simply understanding the structure of the definable sets, say in specific classes containing examples of interest, can already involve deep mathematics*. For instance, Tarski's proof of quantifier elimination for the reals and the cell decomposition theorem for o-minimal structures.

In the examples given so far, as is often the case in mathematics, the specific role of the infinite may be mainly as a kind of foil reflecting the fine structure of compactness, irrespective of the otherwise depth of proofs.

To see the interaction of model theory and set theory which we invoked at the beginning, we need to go further up and further in. (As an aside, already in Hilbert's remarks, via Church, there is an implicit parallel between the understanding of infinite sizes and the development of different models.[6]) Suppose we step back and study the class of all theories.

A thesis of Shelah's groundbreaking *Classification Theory* (1978) is that one can find dividing lines among the class of first-order theories. A dividing line marks a sea change in the combinatorial structure. (The assertion that something is a dividing line requires evidence on both sides: showing that models of theories on one side are all complex in

[5] A clever example is Ax's proof that any injective polynomial map from $\mathbb{C}^n$ to $\mathbb{C}^n$ is surjective. For each finite k and n, there is a sentence $\varphi_{n,k}$ of first order logic in the language $\{+, \times, 0, 1\}$ asserting that any injective map given by $(x_1, \dots, x_n) \mapsto (p_1(x_1, \dots, x_n), \dots, p_n(x_1, \dots, x_n))$ where the p_i are of degree $\leq k$ is surjective. For each prime p, we may write $\bar{F}_p$ as the union of an increasing chain of finite fields. The assertion $\varphi_{n,k}$ is true in each finite field because it is finite, and it follows (e.g. from its logical form) that $\varphi_{n,k}$ holds in $\bar{F}_p$. For $\mathfrak{D}$ any nonprincipal ultrafilter on the primes, $\prod_{p \in P} \bar{F}_p / \mathfrak{D}$ is isomorphic to $\mathbb{C}$, so by Łos' theorem $\varphi_{n,k}$ is true in $\mathbb{C}$.

[6] "Hilbert does not say that the order in which the [list of 23] problems are numbered gauges their relative importance, and it is not meant to suggest that he intended this. But he does mention the arithmetical formulation of the concept of the continuum and the discovery of non-Euclidean geometry as being the outstanding mathematical achievements of the preceding century, and gives this as a reason for putting problems in these areas first, " Church [1968].

some given sense, while models of theories on the other side admit some kind of structure theory.) A priori it is not at all obvious that these should exist. Why wouldn't the seemingly unconstrained range and complexity of theories allow for some kind of continuous gradation along any reasonable axis? The example of extremal combinatorics may give a hint: graphs of a given large finite size are not so easily classifiable, but by examining asymptotic growth rates of certain phenomena, jumps may appear.

Stability, the dividing line which has most profoundly influenced the present field, arises in Shelah [1978] from counting limit points. For a theory T and an infinite cardinal λ, we say T is *λ-stable* if for every model M of T of size λ, $|\mathbf{S}(M)| = \lambda$. Conversely, if some model M of T of size λ has $|\mathbf{S}(M)| > \lambda$, T is *λ-unstable*. For a given λ, T is either λ-stable or λ-unstable by definition. But varying λ, the gap appears[7]:

Theorem 3.1 (Shelah 1978)**.** *Any theory T is either* stable*, meaning stable in all λ such that $\lambda^{|T|} = \lambda$, or* unstable*, meaning unstable in all λ.*

This theorem materializes in step with the development of the internal structure theory. The set theoretic scaffolding is not only in the statement, but intricately connected to its development. A few examples from chs. II-III of Shelah [ibid.] will give a flavor:

i) it turns out stability is local: if T is unstable, then there is a single formula φ such that in all λ, we can already get many types just using definable sets which are instances of φ. (This leads to discovering instability has a characteristic combinatorial configuration, the order property.) Its proof is a counting argument relying on the fact that T is unstable in some λ such that $\lambda = \lambda^{|T|}$.

ii) a characteristic property of stable theories is that once there is enough information, types have unique generic extensions to types over larger sets. This is first explained by the finite equivalence relation theorem which studies types over sets A of size at least 2 in models which are $(|A|^{|T|})^+$-saturated.

iii) conversely, large types are essentially controlled by their restrictions to "small" sets. The cardinal defining this use of "small," $\kappa(T)$, is $\leq |T|^+$. Above this cardinal the mist clears and stability's effects are easier to see; one can e.g. characterize larger saturated models of stable theories as those which are $\kappa(T)$-saturated and every maximal indiscernible set has the cardinality of the model.

iv) instability has a more random form and a more rigid form, and at least one must occur. To see the difference between the two by counting types, one has to know that 2^λ can potentially be different from the number of cuts in a dense linear order of size λ (or the number of branches in a tree with λ nodes), which was originally noted by appeal to an independence result.

[7] $|T|$, the size of T, will be the maximum of the size of the language and $\aleph_0$.

Of course, to say infinite cardinals are strongly connected to the development of stability doesn't mean they are necessarily there at the end. The order property, definability of types, forking, the independence property, and the strict order property, to name a few, don't bear the imprint of their origin. This translation is part of model theory's power.

Looking forwards: the effect of stability, inside and outside model theory, has been significant. Despite this conclusive evidence that some dividing lines do exist, and that they can be very useful, further ones have been challenging to find. We know very few in the vast territory of unstable theories, found – like stability – one by one in response to specific counting problems. To go further, perhaps we can try to shift the way in which set theory sounds out model theoretic information.

The reader may wonder: is model theory being described as a kind of extension to the infinite setting of extremal arguments in combinatorics, with the hierarchy of infinite cardinals replacing the natural numbers? This analogy is challenging, but incomplete. It is incomplete because the finitary, extremal picture doesn't seem to provide a precedent or explanation for the role of model theory, which builds in a remarkable way a bridge between the infinite combinatorial world and a more algebraic one. Still, it is challenging because it leads us to ask what in the infinite setting may play the role of those crucial tools of the combinatorial setting, which may seem to have little place in current model theoretic arguments – namely, probability and randomness.

4 Ultrapowers

Only in the move to ultrapowers does one really recover, albeit in a metaphorical way, that other key ingredient of extremal arguments, the understanding of probability and average behavior.

Stability arises from counting limit points. Recall from 2.3 that saturation is a notion of completeness for a model: λ^+-saturation means all types over all submodels of size at most λ are realized. The ultrapower construction, given formally below, starts with a given model and amplifies it – staying within the elementary class – according to a specific kind of averaging mechanism, an ultrafilter. The resulting larger model, which depends only on the model we began with and the ultrafilter, is called an ultrapower. The level of saturation in the ultrapower reflects whether the given averaging mechanism, applied to the given model, leads to the appearance of many limit points of smaller sets.

If so, this may indicate either that the ultrafilter is powerful, or that the types of the model are not complex. Since we can apply the same ultrafilter to different models and compare the results, however, we can use this construction to compare the 'complexity' of different models (and, dually, of ultrafilters). Restricting to the powerful class of *regular* ultrafilters, whose ability to produce saturation in ultrapowers will be an invariant of the elementary class of the model we begin with, we can use this construction to compare the complexity of theories. Informally for now, the relation on complete countable theories setting $T_1 \trianglelefteq T_2$

> if for any regular ultrafilter $\mathfrak{D}$, if $\mathfrak{D}$-ultrapowers of models of T_2 are sufficiently saturated, so are $\mathfrak{D}$-ultrapowers of models of T_1

is Keisler's order, defined in 1967. It is a pre-order on theories, considered as a partial order on the equivalence classes of theories.

The theorem which convinced the author of this essay, reading around 2005, that it was urgent to study $\trianglelefteq$ further was a theorem in Shelah's *Classification Theory* in a chapter devoted to the ordering. The theorem says that the union of the first two equivalence classes in Keisler's order is precisely the stable theories.

This theorem can be understood as saying that the class of stable theories, which we can see by counting types, can also be seen by asking about good average behavior. Beyond stability, our counting is less useful, and yet the other half, about average behavior, retains its power.

5 In more detail

The idea of a filter was used in 2.2 above to find limit points, but it can also be used to give averages. Maximal filters, called ultrafilters, can be thought of as a coherent choice of which subsets of a given set I are "large."

Definition 5.1. *For an infinite set I, $\mathfrak{D} \subseteq \mathcal{P}(I)$ is an ultrafilter if it is a filter not strictly contained in any other filter.* (*We will assume ultrafilters contain all co-finite sets.*)

The ultraproduct by $\mathfrak{D}$ of a family of models $\langle M_i : i \in I \rangle$ is a model, built in two steps, reflecting the definition of model. First, we define the domain. Identify two elements $\langle a[i] : i \in I \rangle$, $\langle b[i] : i \in I \rangle$ of the Cartesian product $\prod_i M_i$ if $\{i \in I : a[i] = b[i]\} \in \mathfrak{D}$. Definition 2.1 makes this an equivalence relation, and the domain of our ultraproduct N is the set of equivalence classes. Next, fix for transparency a representative of each equivalence class, so that for $a \in N$ and $i \in I$, "$a[i]$" makes sense. The relations, functions, and constants of our language are defined on the ultraproduct by consulting the average of the models: e.g. we say a given k-place relation R holds on $a_1, \dots, a_k$ in N iff $\{i : R(a_1[i], \dots, a_k[i])$ holds in $M_i\} \in \mathfrak{D}$, and for an n-place function f, define $f(a_1, \dots, a_n)$ to be the equivalence class of $\langle b[i] : i \in I \rangle$ where $b[i] = f(a_1[i], \dots, a_n[i])$ computed in M_i. The special case of an ultrapower, where all the factor models are isomorphic, transforms a given structure into a larger, 'amplified' model in the same elementary class. For example, letting $\mathfrak{D}$ be a regular ultrafilter on the set of primes, $\prod_p \overline{\mathbb{F}}_p/\mathfrak{D} \cong \mathbb{C}$, but if we consider the ultrapower $M^P/\mathfrak{D}$ where M is the algebraic closure of the rationals, we also get $\mathbb{C}$.

There is a veiled interaction between the two model theoretic uses of filters: the realization of *types* in the ultrapower, and the *ultrafilter* used in the construction. This is most useful when the ultrafilter is *regular*, ensuring that saturation depends only on finitary input from each factor model. For each regular $\mathfrak{D}$, whether or not an ultrapower $M^I/\mathfrak{D}$ is $|I|^+$-saturated is an invariant of the elementary class of the model M we began with.[8]

Keisler's suggestion was that this could be used to compare theories.

Definition 5.2 (Keisler's order, 1967)**.** *Let T_1, T_2 be complete countable theories.*

$$T_1 \trianglelefteq T_2$$

if for every infinite λ, every regular ultrafilter $\mathfrak{D}$ on λ, every model M_1 of T_1 and every model M_2 of T_2, if $(M_2)^\lambda/\mathfrak{D}$ is λ^+-saturated, then $(M_1)^\lambda/\mathfrak{D}$ is λ^+-saturated.

Informally, say "$\mathfrak{D}$ saturates T" if for some (by regularity of $\mathfrak{D}$, the choice does not matter) model M of T, $M^\lambda/\mathfrak{D}$ is λ^+-saturated: all limit points over small submodels appear. Then Keisler's order puts T_1 less than T_2 if every regular ultrafilter able to saturate T_2 is able to saturate T_1. Note that any two theories may in principle be compared – algebraically closed fields of fixed characteristic, random graphs, real closed fields. Keisler proved his order was well defined and had a minimum and a maximum class (he gave a sufficient condition for membership in each), and asked about its structure.

The crucial chapter on Keisler's order in Shelah [1978] was already mentioned. Its structure on the unstable, non-maximal theories was left there as an important open question. Following Shelah [ibid.], work on Keisler's order stalled for about thirty years. The question was reopened in Malliaris' thesis and in the series of papers Malliaris [2009, 2010b,a, 2012a,b], guided by the perspective described above. Then in the last few years, a very productive ongoing collaboration of Malliaris and Shelah [2015a, 2014a, 2013a,b, 2016a,b, 2018b]... has advanced things considerably.

6 The current picture

Along this road, what does one find?
(a) First, we do indeed see evidence of dividing lines – many more than previously thought. Much remains to be done in understanding them and in characterizing the structure/nonstructure which come with the assertion of a dividing line, but already their appearance, in a region

[8] That can be taken as a definition of regular; alternatively, $\mathfrak{D}$ is a regular ultrafilter on I if there is a set $\{X_\alpha : \alpha < |I|\} \subseteq \mathfrak{D}$ such that the intersection of any infinitely many of its elements is empty. Regular ultrafilters are easy to build and exist on any infinite set.

of theories thought to be relatively tame, is surprising and exciting. In Malliaris and Shelah [2018b], Malliaris and Shelah prove that Keisler's order has infinitely many classes. The theories which witness these different classes come from higher analogues of the countable triangle-free random graph, originally studied by Hrushovski [2002]: the infinite generic tetrahedron-free three-hypergraph, the infinite generic 4-uniform hypergraph with no complete hypergraph on 5 vertices, and so on. The proof shows they may have very different average behavior, as reflected in their differing sensitivity to a certain degree of calibration in the ultrafilters.

These results build on advances in ultrafilter construction, which allow for a greater use of properties of cardinals, even for ultrafilters in ZFC.

Several incomparable classes are known Malliaris and Shelah [2018a], Ulrich [2017], Malliaris and Shelah [2015b] and it may be that future work will reveal many. Perhaps the way that such averages could be perturbed or distorted, and by extension the structure of dividing lines among unstable theories, will be much finer than what we now see. If so, even independence results could be quite useful model theoretically. These may simply witness that the boolean algebras associated to different theories are essentially different, because they react differently to certain exotic averaging mechanisms, when these appear. The internal theories of each equivalence class, giving an account of what allows for the different reactions, would presumably be, like IP or SOP, absolute.

(b) Second, this line of work has led to some surprising theorems about the finite world. These theorems have the following general form. We know that among theories with infinite models, stability is a dividing line, with models of stable theories admitting a strong structure theory. There is a specific combinatorial configuration, the order property, which (in infinite models) characterizes instability. In an infinite graph, instability for the edge relation would correspond to having arbitrarily large *half-graphs*, that is, for all k having vertices $a_1, \dots, a_k$ and $b_1, \dots, b_k$ with an edge between a_i and b_j iff $i < j$. (Note that there are no assertions made about edges among the a's and among the b's, so in forbidding k-half graphs, we forbid a family of configurations.) The thesis of Malliaris and Shelah [2014b] is essentially that finite graphs with no long half-graphs, called *stable graphs*, behave much better than all finite graphs, in the sense predicted by the infinite case.

Szemerédi's celebrated regularity lemma says, roughly, that for every $\epsilon > 0$ there is $N = N(\epsilon)$ such that any sufficiently large graph may be equitably partitioned into $k \leq N$ pieces such that all but at most ϵk^2 pairs of pieces have the edges between them quite evenly distributed (i.e. are ϵ-regular). The elegant picture of this lemma absorbs the general complexity of graphs in two ways: first, by work of Gowers, N is a very large function of ϵ; second, as noticed independently by a number of researchers, the condition that some pairs of pieces be irregular cannnot be removed, as shown by the example of half-graphs (Komlós and Simonovits [1996]). As a graph theorist, one might expect half-graphs to be just one example of bad behavior, not necessarily unique; but in light of the above, a model theorist may guess that in the absence of long half-graphs one will find structured behavior. The stable regularity lemma of Malliaris and Shelah [2014b] shows that indeed, half-graphs are the only reason for irregular pairs: finite stable graphs admit regular partitions with no irregular pairs and the number of pieces singly-exponential in ϵ.

A second theorem in that paper, a stable Ramsey theorem, proves that for each k there is $c = c(k)$ such that if G is a finite graph with no k-half graphs then G contains a clique or an independent set of size $|G|^c$, much larger than what is predicted by Ramsey's theorem. This meets the prediction of the Erdős-Hajnal conjecture, which says that for any finite graph H there is $c = c(H)$ such that if G contains no induced copies of H then G contains a clique or an independent set of size $|G|^c$. But very few other cases of this conjecture are known. What is the contribution of the infinite here? The infinite version of Ramsey's theorem says that a countably infinite graph contains a countable clique or a countable independent set. Extending this to larger cardinals doesn't get far: Erdős-Rado shows the graph and the homogeneous set may not, in general, increase at the same rate. Model theory, however, refracts this result across different classes of theories, and across dividing lines, and in some classes, such as the stable theories, it behaves differently: the existence of large sets of indiscernibles in stable theories implies, a fortiori, that an infinite stable graph of size κ^+ will have a clique or independent set of size κ^+, much larger than predicted by Erdős–Rado. Once one knows where to look, one can find the analogous phenomenon in the finite case (also for hypergraphs).

The stable Ramsey theorem was applied by Malliaris and Terry [2018] to re-prove a theorem in the combinatorics literature, by re-organizing the proof into cases which take advantage of the stable Ramsey theorem, and thus to obtain better bounds for the original theorem; finitary model-theoretic analysis may be useful even where model theoretic hypotheses are not used in the theorems.

It may seem that these theorems of Malliaris and Shelah [2014b], from the first joint paper of Malliaris and Shelah, could in principle have been discovered earlier, and yet they were not. They belong to the perspective of this program in a deep way. They were motivated by work in Malliaris [2010b] and Malliaris [2010a], directed towards Keisler's order, which first applied Szemerédi regularity to study the complexity of formulas (and showed that the simple theories, of special interest in Keisler's order, were in some sense controlled by stable graphs).

(c) Third, by means of these methods model theory has paid an old debt to set theory and general topology, by solving a seventy-year-old problem about cardinal invariants of the continuum. Two infinite cardinals, $\mathfrak{p}$ and $\mathfrak{t}$, known to be uncountable but no larger than the continuum, are shown in Malliaris and Shelah [2013b], Malliaris and Shelah [2016a] to be unconditionally equal. The proof is model-theoretic, and comes in the context of the solution of an a priori unrelated problem, determining a new sufficient condition for maximality in Keisler's order.

In slightly more detail, in order for a regular ultrafilter to handle the most complex theories, those in the maximum class, it must be in some sense very balanced. Distortions and so to speak imperfections which might pass unnoticed in more robust theories will translate immediately in maximal theories to the omission of types. However, a surprising fact from Shelah [1978] is that what is needed for the theory to be complex is not necessarily that it be expressive. A kind of brittleness or overall rigidity as exemplified by the theory of linear order will also suffice. Remarkably, it turns out even less will suffice: the engine of the proof in Malliaris and Shelah [2016a] is in showing that if the ultrafilter can ensure certain paths through trees have upper bounds, it must be strong enough to produce the needed limit points for any theory. The resulting comparison of theories whose models contain the relevant trees, to models of linear orders, via ultrapowers, turns out to be parallel in a precise sense to the comparison of $\mathfrak{p}$ and $\mathfrak{t}$. It was possible to give a fundamental model-theoretic framework encompassing both problems and so to solve them both. A commentary and an expository account of the proof are Moore [2013], Casey and Malliaris [2017].

Still, a model theoretic necessary condition for maximality remains open.
It has been almost ten years since Malliaris [2009]. Profound questions remain, urgent in their simplicity.

Appendix: on definable sets

By convention, we always assume our language $\mathfrak{L}$ contains a binary relation symbol $=$, and that in every $\mathfrak{L}$-model this symbol is interpreted as equality. Besides the symbols of $\mathfrak{L}$, our alphabet for building formulas includes infinitely many variables along with logical symbols $(,), \wedge, \vee, \neg, \iff, \implies, \forall, \exists$.

For awhile let $M_\star$ denote a model for the language $\mathfrak{L}_\star = \{+, \times, -, 0, 1, <\}$, where $+$, $\times$, $-$ are binary function symbols, $<$ is a binary relation, and $0, 1$ are constants. Let us say the domain of $M_\star$ is $\mathbb{R}$ and the symbols have their usual interpretation.[9]

The *terms* of a language are elaborate names. We define terms by induction. All variables and constant symbols are terms; if f is a k-place function symbol and $t_1, \dots, t_k$ are terms, then $f(t_1, \dots, t_k)$ is a term; and a finite string of allowed symbols is a term iff it can be built in finitely many steps in this way.

[9] Pedantically, the form of the symbols makes no demands on their interpretation (other than the basic conditions in the footnote on p. 102); we could, for example, give a perfectly valid model by interpreting both $+$ and $\times$ by projection onto the first coordinate.

The $\mathfrak{L}_\star$-terms include 0, 1, $1+1$ which we may abbreviate 2, $x+1$, $x \times y$, $(x \times 1) + (y + 0)$, $((x \times x) \times x)$ which we may abbreviate x^3. Using similar abbreviations, and dropping parentheses for readability, $x^5 + 15x^2 + 3x + 5$ is also a term.

A key feature of terms is that *if* we are working in a model, and we are given a term along with instructions of which elements of the model to put in for which, if any, variables in the term (recall that in any model, any constant symbols must already refer to specific elements), then the term will evaluate unambiguously to some other element of the model.

Next, by induction, we define *formulas*. Atomic formulas are assertions that a given relation symbol of our language holds on a given sequence of terms. (In $\mathfrak{L}_\star$, the relation symbols are $=$ and $<$, so these will include $x^3+5x+2 = 0$ and also $5+x+15y^2 > 37-z$.) Atomic formulas are formulas. If φ is a formula, then $\neg(\varphi)$ is a formula. If φ and ψ are formulas, then $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \implies \psi)$, $(\varphi \iff \psi)$ are also formulas. If φ is a formula, and x is a variable, then $(\exists x)\varphi$ and $(\forall x)\varphi$ are formulas. A finite string of allowed symbols is a formula iff it can be built in finitely many steps in this way.

In our example here are some more formulas: $(\forall x)(x + 0 = x)$, $(\exists x)(y + x^2 = z)$. Note an important difference between the two. $(\forall x)(x+0 = x)$ is an assertion which will be true or false in any given model; in our given $M_\star$ it is true. By contrast, $(\exists x)(y+x^2 = z)$ is neither true nor false, since it has two free variables; rather, it has a *solution set*, the pairs (a, b) of elements of M such that $(\exists x)(a + x^2 = b)$. In any model N, the solution sets of formulas with one or more [but always finitely many] free variables are called the *definable sets*. The closure properties of the set of formulas show that for each n, the definable sets on N^n form a boolean algebra.

The formulas with no free variables are called *sentences*, and the theory of a model N is the set of all sentences which hold in N. The elementary class of N is the class of all other models in the same language with the same theory. The reason a theory may make assertions about definable sets which are meaningful across different models is by referring to their defining formulas.

When a formula has many free variables, it may be useful to look at the restricted solution set we get after specifying that certain of the free variables take certain values in the model. For example, in the formula $xy^2 + zy + w = 0$ with free variables x, y, z, w, we might want to consider the solution set under specific values of x, z and w. Such a solution set is called *definable with parameters*, the specific values being the parameters. We may wish to record their provenance: given a subset A of a model N, the sets definable with parameters from A are called *A-definable sets*. Finally, a word on types. Given M and $A \subseteq M$, the set of formulas with parameters from A and (say) one free variable can be made into a boolean algebra once the formulas are identified up to logical equivalence (equivalently, identified if they define the same set in M). Its Stone space is the set of types of M over A in the sense of 2.2 above, and its compactness as a topological space is explained by the compactness theorem.

Some examples – here, definable means with parameters:

1) in the model $M_\star$ above, the definable sets include the semialgebraic sets (and it is a theorem that they are exactly the semialgebraic sets). Its elementary class is the class of real closed ordered fields.

2) if $\mathcal{L} = \{+, \times, -, 0, 1\}$ and M is the algebraic closure of the rationals on which the symbols have their usual interpretation, the definable sets include (and, in fact, are) the constructible sets. The elementary class of M is the class of all algebraically closed fields of characteristic zero.

3) if $\mathcal{L} = \{<,\}$, and M is the rationals on which $<$ has its usual interpretation, the definable sets in one free variable are finite unions of points and intervals (and the definable sets in $k > 1$ free variables satisfy a cell decomposition theorem). The elementary class of M, the class of dense linear orders without a first or last element, has only one countable model, up to isomorphism, but 2^λ nonisomorphic models of each uncountable size λ.

Acknowledgments

B. Mazur, A. Peretz, and S. Shelah made helpful comments. The Institute for Advanced Study contributed some wonderful working conditions. Partial support of the National Science Foundation, and of a Minerva Research membership was noted above. Thank you.

References

J. T. Baldwin and A. H. Lachlan (1971). "On strongly minimal sets". *J. Symbolic Logic* 36, pp. 79–96. MR: 0286642.

D. Casey and M. Malliaris (Sept. 2017). "Notes on Cofinality Spectrum Problems". arXiv: 1709.02408 (cit. on p. 112).

A. Church (1968). "Paul J. Cohen and the continuum problem". In: *Proc. Internat. Congr. Math. (Moscow, 1966)*. Izdat. "Mir", Moscow, pp. 15–20. MR: 0232675 (cit. on p. 105).

E. Hrushovski (1990). "Unidimensional theories are superstable". *Ann. Pure Appl. Logic* 50.2, pp. 117–138. MR: 1081816.

– (2002). "Pseudo-finite fields and related structures". In: *Model theory and applications*. Vol. 11. Quad. Mat. Aracne, Rome, pp. 151–212. MR: 2159717 (cit. on p. 110).

H. J. Keisler (1967). "Ultraproducts which are not saturated". *J. Symbolic Logic* 32, pp. 23–46. MR: 0218224.

J. Komlós and M. Simonovits (1996). "Szemerédi's regularity lemma and its applications in graph theory". In: *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*. Vol. 2. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, pp. 295–352. MR: 1395865 (cit. on p. 111).

J. Łoś (1962). "Remarks on foundations of probability: semantical interpretation of the probability of formulas". In: *Proceedings of the International Congress of Mathematicians*.

M. Malliaris (2009). "Realization of φ-types and Keisler's order". *Ann. Pure Appl. Logic* 157.2-3, pp. 220–224. MR: 2499710 (cit. on pp. 109, 112).

– (2010a). "Edge distribution and density in the characteristic sequence". *Ann. Pure Appl. Logic* 162.1, pp. 1–19. MR: 2720656 (cit. on pp. 109, 111).

– (2010b). "The characteristic sequence of a first-order formula". *J. Symbolic Logic* 75.4, pp. 1415–1440. MR: 2767977 (cit. on pp. 109, 111).

– (2012a). "Hypergraph sequences as a tool for saturation of ultrapowers". *J. Symbolic Logic* 77.1, pp. 195–223. MR: 2951637 (cit. on p. 109).

– (2012b). "Independence, order, and the interaction of ultrafilters and theories". *Ann. Pure Appl. Logic* 163.11, pp. 1580–1595. MR: 2959661 (cit. on p. 109).

M. Malliaris and A. Peretz (2017). "What simplicity is not". In: *Simplicity: ideals of practice in mathematics and the arts*. Math. Cult. Arts. Springer, Cham, pp. 51–58. MR: 3676543.

M. Malliaris and A. Pillay (2016). "The stable regularity lemma revisited". *Proc. Amer. Math. Soc.* 144.4, pp. 1761–1765. MR: 3451251.

M. Malliaris and S. Shelah (2013a). "A dividing line within simple unstable theories". *Adv. Math.* 249, pp. 250–288. MR: 3116572 (cit. on p. 109).

– (2013b). "General topology meets model theory, on $\mathfrak{p}$ and $\mathfrak{t}$". *Proc. Natl. Acad. Sci. USA* 110.33, pp. 13300–13305. MR: 3105597 (cit. on pp. 109, 111).

– (2014a). "Model-theoretic properties of ultrafilters built by independent families of functions". *J. Symb. Log.* 79.1, pp. 103–134. MR: 3226014 (cit. on p. 109).

– (2014b). "Regularity lemmas for stable graphs". *Trans. Amer. Math. Soc.* 366.3, pp. 1551–1585. MR: 3145742 (cit. on pp. 110, 111).

– (2015a). "Constructing regular ultrafilters from a model-theoretic point of view". *Trans. Amer. Math. Soc.* 367.11, pp. 8139–8173. MR: 3391912 (cit. on p. 109).

– (2015b). "Manuscript F1530" (cit. on p. 110).

– (2016a). "Cofinality spectrum theorems in model theory, set theory, and general topology". *J. Amer. Math. Soc.* 29.1, pp. 237–297. MR: 3402699 (cit. on pp. 109, 111, 112).

– (2016b). "Existence of optimal ultrafilters and the fundamental complexity of simple theories". *Adv. Math.* 290, pp. 614–681. MR: 3451934 (cit. on p. 109).

– (Sept. 2017). "A new look at interpretability and saturation". arXiv: 1709.04899.

M. Malliaris and S. Shelah (Sept. 2018a). “An example of a new simple theory”. arXiv: 1804.03254 (cit. on p. 110).

– (2018b). “Keisler’s order has infinitely many classes”. Israel J. Math. (cit. on pp. 109, 110).

M. Malliaris and C. Terry (2018). “On unavoidable induced subgraphs in large prime graphs”. *Journal of Graph Theory* 88.2, pp. 255–270 (cit. on p. 111).

M. Mashaal (2006). *Bourbaki: A secret society of mathematics*. American Mathematical Society (cit. on p. 103).

J. T. Moore (2013). “Model theory and the cardinal numbers $\mathfrak{p}$ and $\mathfrak{t}$”. *Proc. Natl. Acad. Sci. USA* 110.33, pp. 13238–13239. MR: 3105596 (cit. on p. 112).

M. Morley (1965). “Categoricity in power”. *Trans. Amer. Math. Soc.* 114, pp. 514–538. MR: 0175782.

S. Shelah (1978). *Classification theory and the number of nonisomorphic models*. Vol. 92. Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Co., Amsterdam-New York, pp. xvi+544. MR: 513226 (cit. on pp. 106, 109, 112).

E. Szemerédi (1975). “On sets of integers containing no k elements in arithmetic progression”. *Acta Arith.* 27. Collection of articles in memory of JuriĭVladimirovič Linnik, pp. 199–245. MR: 0369312.

D. Ulrich (2017). “Keisler’s Order is Not Linear, Assuming a Supercompact”. To appear, J. Symb. Logic. arXiv: 1702.01237 (cit. on p. 110).

B. I. Zilber (1984). “Strongly minimal countably categorical theories. III”. *Sibirsk. Mat. Zh.* 25.4, pp. 63–77. MR: 754741.

Received 2017-12-05.

Maryanthe Malliaris
Department of Mathematics
University of Chicago
5734 S. University Avenue
Chicago, IL 60637
USA
mem@math.uchicago.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (117–142)

QUIVERS WITH RELATIONS FOR SYMMETRIZABLE CARTAN MATRICES AND ALGEBRAIC LIE THEORY

Christof Geiß

Abstract

We give an overview of our effort to introduce (dual) semicanonical bases in the setting of symmetrizable Cartan matrices.

1 Introduction

One of the original motivations of Fomin and Zelevinsky for introducing cluster algebras was "to understand, in a concrete and combinatorial way, G. Lusztig's theory of total positivity and canonical bases" Fomin [2010]. This raised the question of finding a cluster algebra structure on the coordinate ring of a unipotent cell, and to study its relation with Lusztig's bases. In a series of works culminating with Geiß, Leclerc, and Schröer [2011] and Geiß, Leclerc, and Schröer [2012], we showed that the coordinate ring of a unipotent cell of a symmetric Kac-Moody group has indeed a cluster algebra structure, whose cluster monomials belong to the dual of Lusztig's semicanonical basis of the enveloping algebra of the attached Kac-Moody algebra. Since the semicanonical basis is built in terms of constructible functions on the complex varieties of nilpotent representations of the preprojective algebra of a quiver, it is not straightforward to extend those results to the setting of symmetrizable Cartan matrices, which appears more natural from the Lie theoretic point of view. The purpose of these notes is to give an overview of Geiß, Leclerc, and Schröer [2017a] - Geiß, Leclerc, and Schröer [2017d], where we are trying to make progress into this direction.

The starting point of our project was Hernandez and Leclerc [2016], where they observed that certain quivers with potential allowed to encode the q-characters of the Kirillov-Reshetikhin modules of the quantum loop algebra $U_q(L\mathfrak{g})$, where $\mathfrak{g}$ is a complex simple Lie algebra of arbitrary Dynkin type. This quiver with potential served as model for the definition of our generalized preprojective algebras $\Pi = \Pi_K(C, D)$ associated to a symmetrizable Cartan matrix C with symmetrizer D over an arbitrary field K, which extends

MSC2010: primary 16G20; secondary 17B67, 13F60, 14M15.

the classical construction of Gelfand and Ponomarev [1979]. After the completion of a preliminary version of Geiß, Leclerc, and Schröer [2017a] we learned that Cecotti and Del Zotto [2012] and Yamakawa [2010] had introduced similar constructions for quite different reasons. In comparison to the classical constructions of Dlab and Ringel [1974], Dlab and Ringel [1980] for a symmetrizable Cartan matrix C, we replace field extensions by truncated polynomial rings. Many of the core results of representations of species carry over over to this setting if we restrict our attention to the so-called locally free modules, see Geiß, Leclerc, and Schröer [2017a]. In particular, we have for each orientation Ω of C an algebra $H = H_K(C, D, \Omega)$ such that in many respects Π can be considered as the preprojective algebra of H. Our presentation of these results in Section 3 is inspired by the thesis Geuenich [2016], he obtains similar results for a larger class of algebras.

Since our construction works in particular over algebraically closed fields, we can extend to our algebras H and Π several basic results about representation varieties of quivers and of varieties of nilpotent representations of the preprojective algebra of a quiver in our new context, again if we restrict our attention to locally free modules, see Section 4. Nandakumar and Tingley [2016] obtained similar results by studying the set of K-rational points of the representation scheme of a species preprojective algebra, which is defined over certain infinite, non algebraically closed fields K.

In our setting we can take $K = \mathbb{C}$, and study algebras of constructible functions on those varieties of locally free modules and realize in this manner the universal enveloping algebra $U(\mathfrak{n})$ of the positive part $\mathfrak{n}$ of a complex semisimple Lie algebra, together with a Ringel type PBW-basis in terms of the representations of H. For arbitrary symmetrizable Cartan matrices we can realize $U(\mathfrak{n})$ together with a semicanonical basis, modulo our *support conjecture*, see Section 5.

Conventions. We use basic concepts from representation theory of finite dimensional algebras, like Auslander-Reiten theory or tilting theory without further reference. A good source for this material is Ringel [1984]. For us, a quiver is an oriented graph $Q = (Q_0, Q_1, s, t)$ with vertex set Q_0, arrow set Q_1 and functions $s, t\colon Q_1 \to Q_0$ indicating the start and terminal point of each arrow. We also write $D = \operatorname{Hom}_K(-, K)$. We say that an A-module M is *rigid* if $\operatorname{Ext}^1_A(M, M) = 0$.

2 Combinatorics of symmetrizable Cartan matrices

2.1 Symmetrizable Cartan matrices and quivers. Let $I = \{1, 2, \dots, n\}$. A *symmetrizable Cartan matrix* is an integer matrix $C = (c_{ij}) \in \mathbb{Z}^{I\times I}$ such that the following holds:

- $c_{ii} = 2$ for all $i \in I$ and $c_{ij} \leq 0$ for all $i \neq j$,

- there exist $(c_i)_{i\in I} \in \mathbb{N}_+^I$ such that $\operatorname{diag}(c_1, \dots, c_n) \cdot C$ is a symmetric.

In this situation $D := \operatorname{diag}(c_1, \dots, c_n) \in \mathbb{Z}^{I\times I}$ is called the *symmetrizer* of C. Note that the symmetrizer is not unique. In particular, for all $k \in \mathbb{N}_+$ also kD is a symmetrizer of C.

It is easy to see that the datum (C, D) of a symmetrizable Cartan matrix C and its symmetrizer D is equivalent to displaying a weighted graph $(\Gamma, \underline{d})$ with

- I the set of vertices of Γ,
- $g_{ij} := \gcd(c_{ij}, c_{ji})$ edges between i and j,
- $\underline{d}\colon I \to \mathbb{N}_+, i \mapsto c_i$.

Here we agree that $\gcd(0,0) = 0$. We have then $c_{ij} = -\frac{\operatorname{lcm}(c_i,c_j)}{c_i} g_{ij}$ for all $i \neq j$.

2.2 Bilinear forms, reflections and roots. We identify the root lattice of the Kac-Moody Lie algebra $\mathfrak{g}(C)$ associated to C with $\mathbb{Z}^I = \oplus_{i\in I}\mathbb{Z}\alpha_i$, where the simple roots $(\alpha_i)_{i\in I}$ form the standard basis. We define on $\mathbb{Z}^I$ by

$$(\alpha_i, \alpha_j)_{C,D} = c_i c_{ij},$$

a symmetric bilinear form. The *Weyl group* $W = W(C)$ is the subgroup of $\operatorname{Aut}(\mathbb{Z}^I)$, which is generated by the simple reflections s_i for $i \in I$, where

$$s_i(\alpha_j) = \alpha_j - c_{ij}\alpha_i.$$

The *real roots* are the set

$$\Delta_{\text{re}}(C) := \cup_{i\in I} W(\alpha_i).$$

The *fundamental region* is

$$F := \{\alpha \in \mathbb{N}^I \mid \operatorname{supp}(\alpha) \text{ is connected, and } (\alpha, \alpha_i)_{C,D} \leq 0 \text{ for all } i \in I\}.$$

Here, $\operatorname{supp}(\alpha)$ is the full subgraph of $\Gamma(C)$ with vertex set $\{i \in I \mid \alpha(i) \neq 0\}$. Then the imaginary roots are by definition the set

$$\Delta_{\text{im}}(C) := W(F) \cup W(-F).$$

Finally the set of all roots is

$$\Delta(C) := \Delta_{\text{re}} \cup \Delta_{\text{im}}(C).$$

The positive roots are $\Delta^+(C) := \Delta(C) \cap \mathbb{N}^I$, and it is remarkable that $\Delta(C) = \Delta^+(C) \cup -\Delta^+(C)$.

A sequence $\mathbf{i} = (i_1, i_2, \dots, i_l) \in I^l$ is called a *reduced expression* for $w \in W$ if $w = s_{i_l} \cdots s_{i_2} s_{i_1}$ and w can't be expressed as a product of less than $l = l(w)$ reflections of the form s_i $(i \in I)$. In this case we set

$$\beta_{\mathbf{i},k} := s_{i_1} s_{i_2} \cdots s_{i_{k-1}}(\alpha_{i_k}) \text{ and } \gamma_{\mathbf{i},k} := s_{i_l} s_{i_{l-1}} \cdots s_{i_{k+1}}(\alpha_{i_k}) \tag{2.1}$$

for $k = 1, 2, \dots, l$, and understand $\beta_{\mathbf{i},1} = \alpha_{i_1}$ as well as $\gamma_{\mathbf{i},l} = \alpha_{i_l}$. It is a standard fact that $\beta_{\mathbf{i},k} \in \Delta^+$ for $k = 1, 2, \dots, l$, and that these roots are pairwise different. Obviously,

$$w(\beta_{\mathbf{i},k}) = -\gamma_{\mathbf{i},k} \text{ for } k = 1, 2, \dots, l.$$

The following result is well known.

Proposition 2.1. *For a connected, symmetrizable Cartan matrix C the following are equivalent:*

- *C is of Dynkin type.*
- *The Weyl group $W(C)$ is finite.*
- *The root system $\Delta(C)$ is finite*
- *All roots are real: $\Delta(C) = \Delta_{\mathrm{re}}(C)$.*

Moreover, if in this situation $\mathbf{i}$ is a reduced expression for w_0, the longest element of W, then $\Delta^+ = \{\beta_{\mathbf{i},1}, \beta_{\mathbf{i},2}, \dots, \beta_{\mathbf{i},l}\}$.

2.3 Orientation and Coxeter elements. An *orientation* of C is a set $\Omega \subset I \times I$ such that

- $|\Omega \cap \{(i,j),(j,i)\}| > 0 \iff c_{ij} < 0$,
- for each sequence $i_1, i_2, \dots, i_{k+1}$ with $(i_j, i_{j+1}) \in \Omega$ for $j = 1, 2, \dots, k$ we have $i_1 \neq i_{k+1}$.

The orientation Ω can be interpreted as upgrading the weighted graph $(\Gamma, \underline{d})$ of (C, D) to a weighted quiver $(Q^\circ, \underline{d})$ with g_{ij} arrows $\alpha_{ij}^{(1)}, \dots \alpha_{ij}^{(g_{ij})}$ from j to i if $(i,j) \in \Omega$, such that $Q^\circ = Q^\circ(C, \Omega)$ has no oriented cycles.

For an orientation Ω of the symmetrizable Cartan matrix $C \in \mathbb{Z}^{I \times I}$ and $i \in I$ we define

$$s_i(\Omega) := \{(r,s) \in \Omega \mid i \notin \{r,s\}\} \cap \{(s,r) \in I \times I \mid (r,s) \in \Omega \text{ and } i \in \{r,s\}\}.$$

Thus, in $Q^\circ(C, s_i(\Omega))$ the orientation of precisely the arrows in $Q^\circ(C,\Omega)$, which are incident with i, is changed. If i is a sink or a source of $Q^\circ(C,\Omega)$ then $s_i(\Omega)$ is also an orientation of C. It is convenient to define

$$\Omega(-,i) := \{j \in I \mid (j,i) \in \Omega\} \text{ and } \Omega(j,-) := \{i \in I \mid (j,i) \in \Omega\}.$$

We have on $\mathbb{Z}^I$ the *non-symmetric bilinear form*

$$\langle -,- \rangle_{C,D,\Omega} \colon \mathbb{Z}^I \times \mathbb{Z}^I \to \mathbb{Z}, (\alpha_i, \alpha_j) \mapsto \begin{cases} c_i & \text{if } i = j, \\ c_i c_{ij} & \text{if } (j,i) \in \Omega, \\ 0 & \text{else.} \end{cases} \tag{2.2}$$

We leave it as an exercise to verify that

$$\langle \alpha, \beta \rangle_{C,D,\Omega} = \langle s_i(\alpha), s_i(\beta) \rangle_{C,D,s_i(\Omega)} \tag{2.3}$$

if i is a sink or a source for Ω.

We say that a reduced expression $\mathbf{i} = (i_1, i_2, \dots, i_l)$ of $w \in W$ is $+$-*admissible* for Ω if i_1 is a sink of $Q^\circ(C,\Omega)$, and i_k is a sink of $Q^\circ(C, s_{i_{k-1}} \cdots s_{i_2} s_{i_1}(\Omega))$ for $k = 2, 3, \dots, l$. If moreover $l = n$ and $\{i_1, \dots, i_n\} = I$, we say that $c = s_{i_n} \cdots s_{i_2} s_{i_1}$ is the *Coxeter element* for (C,Ω).

2.4 Kac-Moody Lie algebras. For a symmetrizable Cartan matrix $C \in \mathbb{Z}^{I\times I}$, the derived Kac-Moody Lie algebra $\mathfrak{g}' = \mathfrak{g}'(C)$ over the complex numbers has a presentation by $3n$ generators e_i, h_i, f_i $(i \in I)$ subject to the following relations:

(i) $[e_i, f_j] = \delta_{ij} h_i$;

(ii) $[h_i, h_j] = 0$;

(iii) $[h_i, e_j] = c_{ij} e_j, \quad [h_i, f_j] = -c_{ij} f_j$;

(iv) $(\operatorname{ad} e_i)^{1-c_{ij}}(e_j) = 0, \quad (\operatorname{ad} f_i)^{1-c_{ij}}(f_j) = 0 \qquad (i \neq j)$.

Note that for C of Dynkin type this is the Serre presentation of the corresponding semisimple Lie algebra. In case rank $C < |I|$ we have of $\mathfrak{g}'(C) \neq \mathfrak{g}(C)$ and the latter has in this case a slightly larger Cartan subalgebra, which makes for a more complicated definition, see for example Geiß, Leclerc, and Schröer [2017c, Sec. 5.1] for a few more details. Of course, the main reference is Kac [1990].

Let $\mathfrak{n} = \mathfrak{n}(C)$ be the Lie subalgebra generated by the e_i $(i \in I)$. Then $U(\mathfrak{n})$ is the associative $\mathbb{C}$-algebra with generators e_i $(1 \le i \le n)$ subject to the relations

$$(\operatorname{ad} e_i)^{1-c_{ij}}(e_j) = 0, \qquad (i, j \in I, i \neq j). \tag{2.4}$$

$U(\mathfrak{n})$ is $\mathbb{N}^I$ graded with $\deg(e_i) = \alpha_i$ $(i \in I)$. With

$$\mathfrak{n}_\alpha := \mathfrak{n} \cap U(\mathfrak{n})_\alpha \text{ for } \alpha \in \Delta^+(C)$$

we recover the usual root space decomposition of $\mathfrak{n}$.

3 Quivers with relations for symmetrizable Cartan matrices

We keep the notations from the previous section, in particular $C \in \mathbb{Z}^{I\times I}$ is a symmetrizable Cartan matrix with symmetrizer D and Ω is an orientation for C.

3.1 A class of 1-Iwanaga-Gorenstein algebras. Let K be a field and $Q = Q(C, D, \Omega)$ the quiver obtained from $Q^\circ(C, D, \Omega)$, see Section 2.3, by adding a loop ϵ_i at each vertex $i \in I$. Then $H = H_K(C, D, \Omega)$ is the path algebra KQ modulo the ideal which is generated by the following relations:

- $\epsilon_i^{c_i}$ for all $i \in I$
- $\epsilon_i^{-c_{ji}/g_{ji}} \alpha_{ij}^{(k)} - \alpha_{ij}^{(k)} \epsilon_j^{-c_{ij}/g_{ij}}$ for all $(i, j) \in \Omega$ and $k = 1, 2, \dots g_{ij}$.

Recall that $g_{ij} = g_{ji} = \gcd(c_{ij}, c_{ji})$, thus $-c_{ij}/g_{ij} = \operatorname{lcm}(c_i, c_j)/c_i$.

For $(i, j) \in \Omega$ let $c'_{ij} = c_{ij}/g_{ij}$ and $c'_{ji} = c_{ji}/g_{ij}$. We may consider the following symmetrizable Cartan matrix, symmetrizer and orientation:

$$C^{(i,j)} = \begin{pmatrix} 2 & c'_{ij} \\ c'_{ji} & 2 \end{pmatrix}, \quad D^{(i,j)} = \begin{pmatrix} c_i & 0 \\ 0 & c_j \end{pmatrix} \quad \text{and} \quad \Omega^{(i,j)} = \{(i, j)\}.$$

Thus,

$$Q^{(i,j)} := Q(C^{(i,j)}, \Omega^{(i,j)}) \quad = \quad \epsilon_i \circlearrowright i \xleftarrow{\alpha_{ij}} j \circlearrowleft \epsilon_j$$

and

$$H^{(i,j)} := H_K(C^{(i,j)}, D^{(i,j)}, \Omega^{(i,j)}) = KQ^{(i,j)}/\langle \epsilon_i^{c_i}, \epsilon_j^{c_j}, \epsilon_i^{-c'_{ji}} \alpha_{ij} - \alpha_{ij} \epsilon_j^{-c'_{ij}} \rangle.$$

Note, that with

$${}_iH'_j := e_i H^{(i,j)} e_j \quad \text{and} \quad H_i = e_i H_i e_i = K[\epsilon_i]/(\epsilon_i^{c_i})$$

it is easy to see that ${}_iH_j := {}_iH_j'^{\oplus g_{ij}}$ is a H_i-H_j-bimodule, which is free of rank $-c_{ij}$ as a H_i-module, and free of rank $-c_{ji}$ as H_j-(right)-module. If we define similarly $H^{(j,i)} :=$

$H_K(C^{(i,j)}, D^{(i,j)}, \{(j,i)\})$ and ${}_jH_i' := e_j H^{(j,i)} e_j$, then ${}_jH_i = {}_jH_i'^{\oplus g_{ij}}$ is a H_j-H_i-bimodule, which is free of rank $-c_{ji}$ as H_j-module and free of rank $-c_{ij}$ as H_i-(left)-module. It is easy to see that we get an isomorphism of H_i-H_j-bimodules

$$ {}_iH_j \cong \operatorname{Hom}_K({}_jH_i, K). $$

The adjunction yields for H_k-modules M_k, for $k \in \{i, j\}$, a natural isomorphism of vector spaces

$$ \operatorname{Hom}_{H_i}({}_iH_j \otimes_{H_j} M_j, M_i) \to \operatorname{Hom}_{H_j}(M_j, {}_jH_i \otimes_{H_i} M_i),\ f \mapsto f^{\vee}. \tag{3.1} $$

Quite similarly to the representation theory of modulated graphs, in the sense of Dlab and Ringel [1974], we have the following basic results from Geiß, Leclerc, and Schröer [2017a, Prop. 6.4] and Geiß, Leclerc, and Schröer [ibid., Prop. 7.1].

Proposition 3.1. *Set $H := H_K(C, D, \Omega)$. With $S := \times_{i\in I} H_i$ we can consider $B := \bigoplus_{(i,j)\in\Omega} {}_iH_j$ as an S-S-bimodule and find:*

(a) $H \cong T_S(B) := \bigoplus_{j\in\mathbb{N}} B^{\otimes_S j}$, i.e. H is a tensor algebra.

(b) There is a canonical short exact sequence of H-H-bimodules

$$ 0 \to H \otimes_S B \otimes_S H \xrightarrow{\delta} H \otimes_S H \xrightarrow{\text{mult}} H \to 0, $$

where $\delta(h_l \otimes b \otimes h_r) = h_l b \otimes h_r - h_l \otimes b h_r$.

Note that the H-H-bimodules $H \otimes_S B \otimes_S H$ and $H \otimes_S H$ are in general only projective as H-left- or right-modules, but not as bimodules. Anyway, the above sequence yields a functorial projective resolution for certain modules which we are going to define now. We say that a H-module M is *locally free* if $e_i M$ is a free H_i-module for all $i \in I$. In this case we define

$$ \underline{\operatorname{rank}}(M) := (\operatorname{rank}_{H_i}(e_i M))_{i\in I}. $$

For example, there is a unique (indecomposable) locally free H-module E_i with $\underline{\operatorname{rank}}(E_i) = \alpha_i$ for each $i \in I$. For later use we define for all $\mathbf{r} \in \mathbb{N}^I$ the module $\mathbf{E}^{\mathbf{r}} := \oplus_{i\in I} E_i^{\mathbf{r}(i)}$, and observe that $\underline{\operatorname{rank}}(\mathbf{E}^{\mathbf{r}}) = \mathbf{r}$. Let us write down the following consequences of Proposition 3.1, see Geiß, Leclerc, and Schröer [ibid., Sec. 3.1] and Geiß, Leclerc, and Schröer [ibid., Cor.7.1].

Corollary 3.2. *For H as above we have:*

(a) *The projective and injective H-modules are locally free. More precisely we have*

$$\underline{\mathrm{rank}}(He_{i_k}) = \beta_{\mathbf{i},k} \quad \text{and} \quad \underline{\mathrm{rank}}(De_{i_k}H) = \gamma_{\mathbf{i},k} \text{ for } k \in I,$$

where $\mathbf{i}$ *is a reduced expression for the Coxeter element of* (C, Ω).

(b) *Each locally free H-module M has a functorial projective resolution*

$$0 \to H \otimes_S B \otimes_S M \xrightarrow{\delta \otimes M} H \otimes_S M \xrightarrow{\mathrm{mult}} M \to 0.$$

Moreover, if M is not locally free, then $\operatorname{proj.dim} M = \infty$.

(c) *H is 1-Iwanaga-Gorenstein, i.e.* $\operatorname{proj.dim}({}_H DH) \leq 1$ *and* $\operatorname{inj.dim}({}_H H) \leq 1$. *Moreover an H-module M is locally free if and only if* $\operatorname{proj.dim}(M) \leq 1$.

It follows that the Ringel (homological) bilinear form descends as the non-symmetric bilinear form (2.2) to the Grothendieck group of locally free modules, where we identify the classes of the generalized simples E_i with the coordinate vector α_i ($i \in I$), see also Geiß, Leclerc, and Schröer [2017a, Prop. 4.1].

Corollary 3.3. *If M and N are locally free H-modules, we have*

$$\dim \mathrm{Hom}_H(M, N) - \dim \mathrm{Ext}^1_H(M, N) = \langle \underline{\mathrm{rank}}(M), \underline{\mathrm{rank}}(N) \rangle_{C,D,\Omega}.$$

By combining Corollary 3.2 with standard results from Auslander-Reiten theory we obtain now the following result.

Corollary 3.4. *Let M be an indecomposable, non projective, locally free H-module such that the Auslander-Reiten translate $\tau_H M$ is locally free. Then*

$$\underline{\mathrm{rank}}(\tau_H M) = c \cdot (\underline{\mathrm{rank}}(M)),$$

where $c = s_{i_n} \cdots s_{i_1}$ is the Coxeter element for (C, Ω). Moreover, if we take $R \in \mathbb{Z}^{I \times I}$, such that $D \cdot R$ is the matrix of $\langle -, - \rangle_{C,D,\Omega}$ with respect to the standard basis, we get $c = -R^{-1}(C - R)$.

This is the K-theoretic shadow of a deeper connection between the Auslander-Reiten translate and reflection functors, which we will discuss in the next subsection.

3.2 Auslander-Reiten theory and Coxeter functors. By Proposition 3.1 we may view $H = H_K(C, D, \Omega)$ as a tensor algebra. Thus, we identify a H-module M naturally with a S-module $\mathbf{M} = \oplus_{i \in I} M_i$ together with an element $(M_{ij})_{(i,j) \in \Omega}$ of

$$H(\mathbf{M}) := \bigoplus_{(i,j) \in \Omega} \in \mathrm{Hom}_{H_i}({}_i H_j \otimes_{H_j} M_j, H_i). \tag{3.2}$$

Write $s_i(H) := H_k(C, D, s_i(\Omega))$ for any $i \in I$. If k is a sink of $Q^\circ(C, \Omega)$, we have for each H-module M a canonical exact sequence
(3.3)

$$0 \to \operatorname{Ker}(M_{k,\text{in}}) \to \bigoplus_{j \in \Omega(k,-)} {}_kH_j \otimes_{H_j} M_j \xrightarrow{M_{k,\text{in}}} M_k, \text{ where } M_{k,\text{in}} = \oplus_{j \in \Omega(k,-)} M_{kj}.$$

We can define now the BGP-reflection functor

$$F_k^+ \colon \operatorname{rep}(H) \to \operatorname{rep}(s_i(H)), \quad (F_k^+ M)_i = \begin{cases} M_i & \text{if } i \neq k, \\ \operatorname{Ker}(M_{k,\text{in}}) & \text{if } i = k. \end{cases}$$

We can moreover define in this situation dually the left adjoint $F_k^- \colon \operatorname{rep}(s_k(H)) \to \operatorname{rep}(H)$. Note that k is a source of $Q^\circ(C, s_k\Omega)$. See Geiß, Leclerc, and Schröer [ibid., Sec. 9.2] for more details. We observe that the definitions imply easily the following:

Lemma 3.5. *If k is a sink for Ω and M is a locally free H-module which has no direct summand isomorphic to E_k and $F_k^+(M)$ is locally free, then $\underline{\operatorname{rank}}(F_k^+ M) = s_k(\underline{\operatorname{rank}}(M))$.*

The proof of Geiß, Leclerc, and Schröer [ibid., Prop. 9.6] implies the following, less obvious result:

Lemma 3.6. *Suppose that k is a sink for Ω and M a locally free rigid H-module, with no direct summand isomorphic to E_k, then* $\operatorname{Hom}_H(M, E_k) = 0$.

We can interpret F_k^+ as a kind of APR-tilting functor Auslander, Platzeck, and Reiten [1979]. See Geiß, Leclerc, and Schröer [2017a, Sec. 9.3] for a proof of this non-trivial result.

Theorem 3.7. *Let k be a sink of $Q^\circ(C, \Omega)$. Then $X := {}_HH/He_k \oplus \tau^- He_k$ is a classical tilting module for H. With $B := \operatorname{End}_H(X)^{\text{op}}$ we have an equivalence $S \colon \operatorname{rep}(s_k(H)) \to \operatorname{rep}(B)$ such that the functors $S \circ F_l^+$ and $\operatorname{Hom}_H(X, -)$ are isomorphic.*

Standard tilting theory arguments and Auslander–Reiten theory, together with Lemma 3.5 and Lemma 3.6 yield the following important consequence:

Corollary 3.8. *Let $k \in I$ be a sink for Ω and M a locally free rigid H-module, then $F_k^+(M)$ is a rigid, locally free $s_k(H)$-module.*

Consider the algebra automorphism of H, which is defined by multiplying the non-loop arrows of $Q(C, \Omega)$ by -1. It induces the so called *twist* automorphism $T \colon \operatorname{rep}(H) \to \operatorname{rep}(H)$. Moreover, let $s_{i_n} \cdots s_{i_2} s_{i_1}$ be the Coxeter element for (C, Ω), corresponding to the $+$-admissible sequence $i_1, i_2, \ldots, i_n$, see Section 2.3. Now we can define the *Coxeter functor*

$$C^+ := F_{i_n}^+ \circ \cdots \circ F_{i_2}^+ \circ F_{i_1}^+ \colon \operatorname{rep}(H) \to \operatorname{rep}(H).$$

Following ideas of Gabriel [1980, Sec. 5], by a careful comparison of the definitions of the reflection functors and Auslander–Reiten translate, we obtain the following result. See Geiß, Leclerc, and Schröer [2017a, Sec. 10] for the lengthy proof.

Theorem 3.9. *With the H-H-bimodule $Y := \operatorname{Ext}^1_H(DH, H)$ we have an isomorphism of endofunctors of* $\operatorname{rep}(H)$*:*

$$\operatorname{Hom}_H(Y, -) \cong T \circ C^+$$

If M is locally free, we have functorial isomorphisms

$$\tau_H(M) \cong \operatorname{Hom}_H(Y, M) \quad \text{and} \quad \tau_H^- M \cong Y \otimes_H M.$$

In particular, in this case the Coxeter functor C^+ and the Auslander–Reiten translate τ may be identified up to the twist T.

It is not true in general that the Auslander–Reiten translate of a locally free H-module is again locally free. In Geiß, Leclerc, and Schröer [ibid., pp. 13.6–13.8] several examples of this behavior are documented. This motivates the following definition. A H-module M is *τ-locally free* if $\tau^k M$ is locally free for all $k \in \mathbb{Z}$. In particular, rigid locally free modules are τ-locally free. We call an indecomposable H-module *preprojective*, resp. *preinjective*, if it is of the form $\tau^{-k}(He_i)$ resp. $\tau^k(De_i H)$ for some $k \in \mathbb{N}_0$ and $i \in I$. Thus, these modules are particular cases of rigid τ-locally free modules.

3.3 Dynkin type. By combining the findings of previous section with standard Auslander–Reiten theory and the characterization of Dynkin diagrams in Proposition 2.1, we obtain the following analog of Gabriel's theorem, see Geiß, Leclerc, and Schröer [ibid., Thm. 11.10].

Theorem 3.10. *Let $H = H_K(C, D, \Omega)$ be as above. There are only finitely many isomorphism classes of indecomposable, τ-locally free H-modules if and only if C is of Dynkin type. In this case the map $M \mapsto \underline{\operatorname{rank}}(M)$ induces a bijection between the isomorphism classes of indecomposable, τ-locally free modules and the positive roots $\Delta^+(C)$. Moreover, all these modules are preprojective and preinjective.*

Note however, that even for C of Dynkin type, the algebra $H(C, D, \Omega)$ is in most cases not of finite representation type, see Geiß, Leclerc, and Schröer [ibid., Prop. 13.1] for details.

Let C be a symmetrizable Cartan matrix of Dynkin type and $\mathbf{i} = (i_1, i_2, \ldots, i_r)$ a reduced expression for the longest element w_0 of the Weyl group W, which is $+$-admissible for the orientation Ω. With the notation of (2.1) we abbreviate $\beta_j = \beta_{\mathbf{i},j}$ for $j = 1, \ldots, r$, and recall that this gives a complete list of the positive roots. By Theorem 3.10 we have for each j a unique, locally free, indecomposable and rigid representation $M(\beta_j)$ with $\underline{\operatorname{rank}}(M(\beta_j)) = \beta_j$.

Proposition 3.11. *With the above notations we have*

$$\langle \beta_i, \beta_j \rangle_{C,D,\Omega} = \begin{cases} \dim \operatorname{Hom}_H(M(\beta_i), M(\beta_j)) & \text{if } i \leq j, \\ -\dim \operatorname{Ext}^1_H(M(\beta_i), M(\beta_j)) & \text{if } i > j. \end{cases}$$

In particular, $\operatorname{Hom}_H(M(\beta_i), M(\beta_j)) = 0$ *if* $i > j$ *and* $\operatorname{Ext}^1_H(M(\beta_i), M(\beta_j)) = 0$ *if* $i \leq j$.

In fact, by Theorem 3.7 and equation (2.3) we may assume that either $i = 1$ or $j = 1$. In any case $M(\beta_1) = E_{i_1}$ is projective. In the first case we have $\operatorname{Ext}^1_H(E_1, M(\beta_j)) = 0$. In the second case we have $\operatorname{Hom}_H(M(\beta_i), E_{i_1}) = 0$ by Lemma 3.6. Now our claim follows by Corollary 3.3.

The next result is an easy adaptation of similar results by Dlab and Ringel [1979] for species. The proof uses heavily Proposition 3.11 and reflection functors. This version was worked out in Omlor's Masters thesis Omlor [2016], see also Geiß, Leclerc, and Schröer [2017d, Sec. 5].

Proposition 3.12. *With the same setup as above let* $k \in \{1, 2, \ldots, r\}$ *and* $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$ *such that* $\beta_k = \sum_{j=1}^r m_j \beta_j$ *and* $m_k = 0$. *Then* $M(\beta_k)$ *admits a non-trivial filtration by locally free submodules*

$$0 = M_{(0)} \subset M_{(1)} \subset \cdots \subset M_{(r)} = M(\beta_k)$$

such that $M_{(j)}/M_{(j-1)} \cong M(\beta_j)^{m_j}$ *for* $j = 1, 2, \ldots, r$. *It follows, that* $M(\beta_k)$ *has* <u>no</u> *filtration by locally free submodules*

$$0 = M^{(r)} \subset M^{(r-1)} \subset \cdots \subset M^{(0)} = M(\beta_k),$$

such that $\underline{\operatorname{rank}}(M^{(j-1)}/M^{(j)}) = m_j \beta_j$ *for* $j = 1, 2, \ldots, r$.

3.4 Generalized preprojective algebras. Let $\overline{Q} = \overline{Q}(C)$ be the quiver which is obtained from $Q(C, \Omega)$ by inserting for each $(i, j) \in \Omega$ additional g_{ij} arrows $\alpha_{ji}^{(1)}, \ldots, \alpha_{ji}^{(g_{ij})}$ from i to j, and consider the potential

$$W = \sum_{(i,j)\in\Omega} \sum_{k=1}^{g_{ij}} (\alpha_{ji}^{(k)} \alpha_{ij}^{(k)} \epsilon_j^{-c_{ij}/g_{ij}} - \alpha_{ij}^{(k)} \alpha_{ji}^{(k)} \epsilon_i^{-c_{ji}/g_{ij}}).$$

The choice of Ω only affects the signs of the summands of W. Recall that for a cyclic path $\alpha_1\alpha_2\cdots\alpha_l$ in $\overline{Q}$ by definition

$$\partial_\alpha^{\text{cyc}}(\alpha_1\alpha_2\cdots\alpha_l) := \sum_{i\in\{j\in[1,l] \mid \alpha_j=\alpha\}} \alpha_{i+1}\alpha_{i+2}\cdots\alpha_l\alpha_1\alpha_2\cdots\alpha_{i-1}.$$

The generalized preprojective algebra of H is

$$\Pi = \Pi(Q, D) := K\overline{Q}/\langle \partial^{\mathrm{cyc}}_{\alpha}(W)\,|_{\alpha\in\overline{Q}_1},\ \epsilon_i^{c_i}\,|_{i\in I}\rangle.$$

It is easy to see that Π does not depend on the choice of Ω, up to isomorphism. Notice that for $(i, j) \in \Omega$ we have

$$\partial^{\mathrm{cyc}}_{\alpha^{(k)}_{ji}}(W) = \alpha^{(k)}_{ij}\epsilon_j^{-c_{ij}/g_{ij}} - \epsilon_i^{-c_{ji}/g_{ij}}\alpha^{(k)}_{ij}.$$

It follows, that for any orientation Ω of C we can equip $\Pi_K(C, D)$ with a $\mathbb{N}_0$-grading by assigning each arrow $\alpha^{(k)}_{ji}$ with $(i, j) \in \Omega$ degree 1 and the remaining arrows get degree 0. We write then

$$\Pi_K(C, D) = \bigoplus_{i=0}^{\infty}\Pi(C, D, \Omega)_i,$$

and observe that $\Pi_K(C, D, \Omega)_0 = H_K(C, D, \Omega)$. We obtain from Theorem 3.9 the following alternative description of our generalized preprojective algebra, which justifies its name:

Proposition 3.13. *Let C be a symmetrizable Cartan matrix with symmetrizer D, and Ω an orientation for C. Then, with $H = H_K(C, D, \Omega)$ we have*

$$\Pi(C, D, \Omega)_1 \cong \mathrm{Ext}^1_H(DH, H)$$

as an H-H-bimodule, moreover

$$\Pi(C, D) \cong T_H(\mathrm{Ext}^1_H(DH, H)) \quad \textit{and} \quad {}_H\Pi(C, D) \cong \bigoplus_{i\in I, k\in\mathbb{N}_0} \tau_H^{-k} He_i.$$

Here the first isomorphism is an isomorphism of K-algebras, and the second one of H-modules.

Similarly to Proposition 3.1 we have the following straightforward description of our generalized preprojective algebra as a tensor algebra modulo canonical relations Geiß, Leclerc, and Schröer [2017a, Prop. 6.1], which yields a standard bimodule resolution. See Geiß, Leclerc, and Schröer [ibid., Sec. 12.1] for the proof, where we closely follow Crawley-Boevey and Shaw [2006, Lem. 3.1]. See also Brenner, Butler, and King [2002, Sec. 4].

Proposition 3.14. *Let C be a symmetrizable, connected Cartan matrix and $\Pi := \Pi_K(C, D)$. With $\overline{B} := \oplus_{(i,j)\in\Omega}({}_iH_j \oplus {}_jH_i)$ we have $\Pi \cong T_S(\overline{B})/\langle \partial^{\mathrm{cyc}}_{\epsilon_i}(W)\,|_{i\in I}\rangle$, where we interpret $\partial^{\mathrm{cyc}}_{\epsilon_i}(W) \in \overline{B}\otimes_S \overline{B}$ in the obvious way. We obtain an exact sequence of Π-Π-bimodules*

$$\Pi\otimes_S\Pi \xrightarrow{f} \Pi\otimes_S\overline{B}\otimes_S\Pi \xrightarrow{g} \Pi\otimes_S\Pi \xrightarrow{h} \Pi \to 0, \tag{3.4}$$

where

$$f(e_i \otimes e_i) = \partial_{\epsilon_i}^{\mathrm{cyc}}(W) \otimes e_i + e_i \otimes \partial_{\epsilon_i}^{\mathrm{cyc}}(W), \quad g(e_i \otimes b \otimes e_j) = e_i b \otimes e_j - e_i \otimes b e_j$$

and h is the multiplication map. Moreover $\mathrm{Ker}(f) \cong \mathrm{Hom}_\Pi(D\Pi, \Pi)$ *if C is of Dynkin type, otherwise f is injective.*

We collect below several consequences, which can be found with detailed proofs in Geiß, Leclerc, and Schröer [2017a, Sec. 12.2]. They illustrate that locally free Π-modules behave in many aspects like modules over classical preprojective algebras. Note that part (b) is an extension of Crawley-Boevey's remarkable formula Crawley-Boevey [2000, Lem. 1]

Corollary 3.15. *Let C be a connected, symmetrizable Cartan matrix, and $\Pi = \Pi_K(C, D)$ as above. Moreover, let M and N be locally free Π-modules.*

(a) If N finite-dimensional, we have a functorial isomorphism

$$\mathrm{Ext}^1_\Pi(M, N) \cong D\,\mathrm{Ext}^1_\Pi(N, M).$$

(b) If M and N are finite-dimensional, we have

$$\dim \mathrm{Ext}^1_\Pi(M, N) = \dim \mathrm{Hom}_\Pi(M, N) + \dim \mathrm{Hom}_\Pi(N, M) - (\underline{\mathrm{rank}}(M), \underline{\mathrm{rank}}(N))_{C,D}.$$

(c) If C is not of Dynkin type, $\mathrm{proj.dim}(M) \leq 2$.

(d) If C is of Dynkin type, Π is a finite-dimensional, self-injective algebra and $\mathrm{rep}_{\mathrm{l.f.}}(\Pi)$ *is a 2-Calabi-Yau Frobenius category.*

Similar to Corollary 3.2 (b) the complex (3.4) yields (the beginning of) a functorial projective resolution for all locally free Π-modules. Thus (a), (b) and (c) follow by exploring the symmetry of the above complex. For (d) we note that in this case Π is finite-dimensional and ${}_\Pi\Pi$ is a locally free module by Theorem 3.10 and Proposition 3.13.

4 Representation varieties

4.1 Notation. Let K be now an algebraically closed field. For Q a quiver and $\rho_j \in e_{t_j}(KQ_{\geq 2})e_{s_j}$ for $j = 1, 2, \ldots, l$ we set $A = KQ/\langle \rho_1, \ldots, \rho_l \rangle$. Note, that every finite dimensional basic K-algebra is of this form. We abbreviate $Q_0 = I$ and set for $\mathbf{d} \in \mathbb{N}_0^I$:

$$\mathrm{Rep}(KQ, \mathbf{d}) := \times_{a \in Q_1} \mathrm{Hom}_K(K^{\mathbf{d}(sa)}, K^{\mathbf{d}(ta)}) \quad \text{and} \quad \mathrm{GL}_\mathbf{d} := \times_{i \in I} \mathrm{GL}_{\mathbf{d}(i)}(K).$$

The reductive algebraic group $\mathrm{GL}_\mathbf{d}$ acts on $\mathrm{Rep}(KQ, \mathbf{d})$ by conjugation, and the $\mathrm{GL}_\mathbf{d}$-orbits correspond bijectively to the isoclasses of K-representations of Q. For $M \in$

$\operatorname{Rep}(KQ, \mathbf{d})$ and $\rho \in e_i KQ e_j$ we can define $M(\rho) \in \operatorname{Hom}_K(K^{\mathbf{d}(j)}, K^{\mathbf{d}(i)})$ in a natural way. We have then the $\mathrm{GL}_{\mathbf{d}}$-stable, Zariski closed subset

$$\operatorname{Rep}(A, \mathbf{d}) := \{M \in \operatorname{Rep}(KQ, \mathbf{d}) \mid M(\rho_i) = 0 \text{ for } j = 1, 2, \ldots, l\}.$$

The $\mathrm{GL}_{\mathbf{d}}$-orbits on $\operatorname{Rep}(A, \mathbf{d})$ correspond now to the isoclasses of representations of A with dimension vector $\mathbf{d}$. It is in general a hopeless task to describe the irreducible components of the affine variety $\operatorname{Rep}(A, \mathbf{d})$.

4.2 Varieties of locally free modules for H.

The set of locally free representations of $H = H_K(C, D, \Omega)$ is relatively easy to describe. Clearly, for each locally free $M \in \operatorname{rep}(H)$ we have $\underline{\dim}(M) = D \cdot \underline{\operatorname{rank}}(M)$.

Proposition 4.1. *For $\mathbf{r} \in \mathbb{N}^I$ we have the open subset*

$$\operatorname{Rep}_{\text{l.f.}}(H, \mathbf{r}) := \{M \in \operatorname{rep}(H, D \cdot \mathbf{r}) \mid M \textit{ is locally free}\} \subset \operatorname{Rep}(H, D \cdot \mathbf{r}),$$

which is irreducible and smooth with $\dim \operatorname{rep}_{\text{l.f.}}(H, \mathbf{r}) = \dim \mathrm{GL}_{D \cdot \mathbf{r}} - \frac{1}{2}(\mathbf{r}, \mathbf{r})_{C,D}$.

In fact, it is well known that the modules of projective dimension at most 1 form always an open subset of $\operatorname{rep}(A, \mathbf{d})$. One verifies next that $\operatorname{Rep}_{\text{l.f.}}(H, \mathbf{r})$ is a vector bundle over the $\mathrm{GL}_{D \cdot \mathbf{r}}$-orbit $\mathcal{O}(\oplus_{i \in I} E_i^{\mathbf{r}(i)})$, with the fibers isomorphic to the vector space $H(\mathbf{r}) := H(\mathbf{E}^{\mathbf{r}})$, see (3.2).

This yields the remaining claims. Note that the (usually) non-reductive algebraic group

$$G_{\mathbf{r}} := \times_{i \in I} \mathrm{GL}_{\mathbf{r}(i)}(H_i) = \operatorname{Aut}_S(\oplus_{i \in I} E_i^{\mathbf{r}(i)})$$

acts on the affine space $H(\mathbf{r})$ naturally by conjugation, and the orbits are in bijection with isoclasses of locally free H-modules with rank vector $\mathbf{r}$.

As a consequence, if M and N are rigid, locally free modules with $\underline{\operatorname{rank}}(M) = \underline{\operatorname{rank}}(N)$, then already $M \cong N$, since the orbits of rigid modules are open.

4.3 Varieties of E-filtered modules for Π.

Recall the description of $\Pi_K(C, D)$ in Proposition 3.14. A $\overline{H} := T_S(\overline{B})$-module M is given by a S-module $\mathbf{M} = \oplus_{i \in I} M_i$ such that M_i is a H_i-module for $i \in I$, together with an element $(M_{ij})_{(i,j) \in \overline{\Omega}}$ of

$$\overline{H}(\mathbf{M}) := \bigoplus_{(i,j) \in \overline{\Omega}} \operatorname{Hom}_{H_i}({}_iH_j \otimes_{H_j} M_j, M_i),$$

where $\overline{\Omega} = \Omega \cap \Omega^{\mathrm{op}}$. Extending somewhat (3.3) we set

$$M_{i,\mathrm{in}} := \left(\bigoplus_{j \in \overline{\Omega}(i,-)} \mathrm{sgn}(i,j) M_{ij} \right) : \bigoplus_{j \in \overline{\Omega}(i,-)} {}_iH_j \otimes_{H_j} M_j \to M_i \quad \text{and}$$

$$M_{i,\mathrm{out}} := \left(\prod_{j \in \overline{\Omega}(-,i)} M_{ji}^{\vee} \right) : M_i \to \bigoplus_{j \in \overline{\Omega}(-,i)} {}_iM_j \otimes_{H_j} M_j .$$

We define now for any S-module $\mathbf{M}$, as above, the affine variety

$$\mathrm{Rep}^{\mathrm{fib}}(\Pi, \mathbf{M}) := \{(M_{ij})_{(i,j) \in \overline{\Omega}} \in \overline{H}(\mathbf{M}) \mid M_{k,\mathrm{in}} \circ M_{k,\mathrm{out}} = 0 \text{ for all } k \in I\},$$

and observe that the orbits of the, usually non-reductive, group $\mathrm{Aut}_S(\mathbf{M})$ on $\mathrm{Rep}^{\mathrm{fib}}(\Pi, \mathbf{M})$ correspond to the isoclasses of possible structures of representations of Π on $\mathbf{M}$, since the condition $M_{k,\mathrm{in}} \circ M_{k,\mathrm{out}}$ corresponds to the relation $\partial^{\mathrm{cyc}}_{\epsilon_k}(W)$.

Similarly to the previous section we can define the open subset

$$\mathrm{Rep}_{\mathrm{l.f.}}(\Pi, \mathbf{r}) := \{M \in \mathrm{Rep}(\Pi, D \cdot \mathbf{r}) \mid M \text{ locally free}\} \subset \mathrm{Rep}(\Pi, D \cdot \mathbf{r}),$$

and observe that $\mathrm{Rep}_{\mathrm{l.f.}}(\Pi, \mathbf{r})$ is a fiber bundle over the $\mathrm{GL}_{D \cdot \mathbf{r}}$-orbit $\mathcal{O}(E^{\mathbf{r}})$, with typical fiber $\mathrm{Rep}^{\mathrm{fib}}(\Pi, \mathbf{E}^{\mathbf{r}})$. Finally we define for any projective S-module $\mathbf{M}$ the constructible subset

$$\Pi(\mathbf{M}) = \{(M_{ij})_{(i,j) \in \overline{\Omega}} \in \mathrm{Rep}^{\mathrm{fib}}(\Pi, \mathbf{M}) \mid ((M_{ij})_{ij}, \mathbf{M}) \text{ is } \mathbf{E}\text{-filtered}\}.$$

Here, a Π-module X is $\mathbf{E}$-filtered if it admits a flag of submodules $0 = X_{(0)} \subset X_{(1)} \subset \cdots \subset X_{(l)} = X$, such that for all k we have $X_{(k)}/X_{(k-1)} \cong E_{i_k}$ for some $i_1, i_2, \ldots, i_l \in I$. Note that for C symmetric and D trivial this specializes to Lusztig's notion of a nilpotent representation for the preprojective algebra of a quiver. However, if C is not symmetric even in the Dynkin case there exist finite-dimensional, locally free Π-modules which are not $\mathbf{E}$-filtered, see Geiß, Leclerc, and Schröer [2017c, Sec. 8.2.2] for an example.

We consider $\Pi(\mathbf{r})$ with the Zariski topology and call it by a slight abuse of notation a variety. In any case, it makes sense to speak of the dimension of $\Pi(\mathbf{r})$ and we can consider the set

$$\mathrm{Irr}(\Pi(\mathbf{r}))^{\max}$$

of top-dimensional irreducible components of $\Pi(\mathbf{r})$.

Theorem 4.2. *Let C be a symmetrizable generalized Cartan matrix with symmetrizer D and $H = H_K(C, D, \Omega), \Pi = \Pi_K(C, D)$ for an algebraically closed field K. For the spaces $\Pi(\mathbf{r})$ of $\mathbf{E}$-filtered representations of Π we have* *(a)* $\dim \Pi(\mathbf{r}) = \dim H(\mathbf{r}) = \sum_{i \in I} c_i \mathbf{r}(i)^2 - \frac{1}{2}(\mathbf{r}, \mathbf{r})_{C,D}$ *for all* $\mathbf{r} \in \mathbb{N}^I$.

(b) The set $\mathcal{B} := \coprod_{\mathbf{r} \in \mathbb{N}^I} \operatorname{Irr}(\Pi(\mathbf{r}))^{\max}$ *has a natural structure of a crystal of type* $B_C(-\infty)$ *in the sense of Kashiwara. In particular, we have*

$$|\operatorname{Irr}(\Pi(\mathbf{r}))^{\max}| = \dim U(\mathfrak{n})_{\mathbf{r}},$$

where $U(\mathfrak{n})$ *the universal enveloping algebra of the positive part* $\mathfrak{n}$ *of the Kac-Moody Lie algebra* $\mathfrak{g}(C)$.

We will sketch in the next two sections a proof of these two statements, which are the main result of Geiß, Leclerc, and Schröer [2017c].

4.4 Bundle constructions. The bundle construction in this section is crucial. It is our version Geiß, Leclerc, and Schröer [ibid., Sec. 3] of Lusztig's construction Lusztig [1991, Sec. 12].

For $m \in \mathbb{N}$ we denote by $\mathcal{P}_m$ the set of sequences of integers $\mathbf{p} = (p_1, p_2, \dots, p_t)$ with $m \geq p_1 \geq p_2 \geq \cdots \geq p_t \geq 0$. Obviously $\mathcal{P}_{c_k}$ parametrizes the isoclasses of H_k-modules, and we define $H_k^{\mathbf{p}} = \oplus_{j=1}^t H_k/(\epsilon_k^{p_j})$. For $k \in I$ and $M \in \operatorname{rep}(\Pi)$ we set

$$\operatorname{fac}_k(M) := M_k / \operatorname{Im}(M_{k,\mathrm{in}}) \quad \text{and} \quad \operatorname{sub}_k(M) := \operatorname{Ker}(M_{k,\mathrm{out}}).$$

With this we can define

$$\Pi(\mathbf{M})^{k,\mathbf{p}} = \{M \in \Pi(\mathbf{M}) \mid \operatorname{fac}_k(M) \cong H_k^{\mathbf{p}}\} \text{ and}$$
$$\Pi(\mathbf{M})_{k,\mathbf{p}} = \{M \in \Pi(\mathbf{M}) \mid \operatorname{sub}_k(M) \cong H_k^{\mathbf{p}}\}$$

for $\mathbf{p} \in \mathcal{P}_{c_k}$. We abbreviate $\Pi(\mathbf{M})^{k,m} = \Pi(\mathbf{M})^{k,c_k^m}$. In what follows, we will focus our exposition on the varieties of the form $\Pi(\mathbf{M})^{k,\mathbf{p}}$, however one should be aware that similar statements and constructions hold for the dual versions $\Pi(\mathbf{M})_{k,\mathbf{p}}$.

For an $\mathbf{E}$-filtered representation $M \in \operatorname{rep}(\Pi)$ there exists always a $k \in I$ such that $\operatorname{fac}_k(M)$, viewed as an H_k-module, has a non-trivial free summand. It is also important to observe that $\Pi(\mathbf{M})^{k,0}$ is an open subset of $\Pi(\mathbf{M})$.

Fix now $k \in I$, let $\mathbf{M}$ be a projective S-module and $\mathbf{U}$ be a proper, projective S-submodule of $\mathbf{M}$ with $U_j = M_j$ for all $j \neq k$. Thus, $\mathbf{M}/\mathbf{U} \cong E_k^r$ for some $r \in \mathbb{N}_+$, and can choose a (free) complement T_k, such that $M_k = U_k \oplus T_k$. For partitions $\mathbf{p} = (c_k^r, q_1, q_2, \dots, q_t)$ and $\mathbf{q} = (q_1, \dots, q_t)$ in $\mathcal{P}_{c_k}$ we set moreover $\operatorname{Hom}_S^{\mathrm{inj}}(\mathbf{U}, \mathbf{M}) := \{f \in \operatorname{Hom}_S(U, M) \mid f \text{ injective}\}$, and define

$$Y^{k,\mathbf{p},\mathbf{q}} := \{(U, M, f) \in \Pi(\mathbf{U})^{k,\mathbf{q}} \times \operatorname{Rep}^{\mathrm{fib}}(\Pi, \mathbf{M}) \times \operatorname{Hom}_S^{\mathrm{inj}}(\mathbf{U}, \mathbf{M}) \mid f \in \operatorname{Hom}_\Pi(U, M)\}.$$

Note that for $(U, M, f) \in Y^{k,\mathbf{p},\mathbf{q}}$ we have in fact $M \in \Pi(\mathbf{M})^{k,\mathbf{p}}$, and that the group $\mathrm{Aut}_S(\mathbf{U})$ acts freely on $Y^{k,\mathbf{p},\mathbf{q}}$ via

$$g \cdot (U, M, f) := ((g_i U_{ij}(\mathrm{id} \otimes g_j^{-1}))_{(i,j)\in\overline{\Omega}},\ M,\ g \cdot f^{-1}).$$

Lemma 4.3. *Consider in the above situation the diagram*

$$\begin{array}{ccc} & Y^{k,\mathbf{p},\mathbf{q}} & \\ {}^{p'}\swarrow & & \searrow^{p''} \\ \Pi(\mathbf{U})^{k,\mathbf{q}} \times \mathrm{Hom}_S^{\mathrm{inj}}(\mathbf{U}, \mathbf{M}) & & \Pi(\mathbf{M})^{k,\mathbf{p}} \end{array}$$

with $p'(U, M, f) = (U, f)$ and $p''(U, M, f) = M$. Then the following holds:

(a) p' is a vector bundle of rank m, where

$$m = \sum_{j\in\overline{\Omega}(-,k)} \dim_K \mathrm{Hom}_K(T_k, {}_kM_j \otimes_{H_j} M_j) - \dim_K \mathrm{Hom}_{H_k}(T_k, \mathrm{Im}(U_{k,\mathrm{in}})).$$

(b) p'' is a fiber bundle with smooth irreducible fibers isomorphic to

$$\mathrm{Aut}_S(\mathbf{U}) \times \mathrm{Gr}_{H_k}^{T_k}(H_k^{\mathbf{p}}),$$

where $\mathrm{Gr}_{H_k}^{T_k}(H_k^{\mathbf{p}}) := \mathrm{Hom}_{H_k}^{\mathrm{surj}}(H_k^{\mathbf{p}}, T_k)/\, \mathrm{Aut}_{H_k}(T_k).$

Corollary 4.4. *In the situation of Lemma 4.3, the correspondence*

$$Z' \mapsto p''(p'^{-1}(Z' \times \mathrm{Hom}_S^{\mathrm{inj}}(\mathbf{U}, \mathbf{M})) := Z''$$

induces a bijection between the sets of irreducible components $\mathrm{Irr}(\Pi(U)^{k,\mathbf{q}})$ *and* $\mathrm{Irr}(\Pi(\mathbf{M})^{k,\mathbf{p}})$*. Moreover we have then*

$$\dim Z'' - \dim Z' = \dim H(\mathbf{M}) - \dim(\mathbf{U}).$$

Note, that this implies already part (a) of Theorem 4.2. In fact the Corollary allows us to conclude by induction that $\dim \Pi(\mathbf{r}) \leq \dim \mathrm{Rep}^{\mathrm{fib}}(H, \mathbf{r})$. On the other hand, we can identify $H(\mathbf{r})$ with an irreducible component of $\Pi(\mathbf{r})$.

4.5 Crystals. For $M \in \mathrm{rep}(\Pi)$ and $j \in I$ there are two canonical short exact sequences

$$0 \to K_j(M) \to M \to \mathrm{fac}_j(M) \to 0 \quad \text{and} \quad 0 \to \mathrm{sub}_j(M) \to M \to C_j(M) \to 0.$$

We define recursively that M is a *crystal module* if $\mathrm{fac}_j(M)$ and $\mathrm{sub}_j(M)$ are locally free for all $j \in I$, and $K_j(M)$ as well as $C_j(M)$ are crystal modules for all $j \in I$. Clearly, if M is a crystal module, for all $j \in I$ there exist $\varphi_j(M), \varphi_j^*(M) \in \mathbb{N}$ such that

$$\mathrm{sub}_j(M) \cong E_j^{\varphi_j(M)} \text{and} \quad \mathrm{fac}_j(M) \cong E_j^{\varphi_j^*(M)}. \tag{4.1}$$

Note moreover, that crystal modules are by construction $\mathbf{E}$-filtered. It is now easy to see that for all projective S-modules $\mathbf{M}$ the set

$$\Pi(\mathbf{M})^{\mathrm{cr}} := \{M \in \Pi(\mathbf{M}) \mid M \text{ is a crystal representation}\}$$

is a constructible subset of $\Pi(\mathbf{M})$. The following result from Geiß, Leclerc, and Schröer [2017c, Sec. 4] is crucial for the proof of Theorem 4.2 (b). It has no counterpart for the case of trivial symmetrizers.

Proposition 4.5. *For each projective S-module $\mathbf{M}$ the set $\Pi(\mathbf{M})^{\mathrm{cr}}$ is a dense and equidimensional subset of the union of all top dimensional irreducible components of $\Pi(\mathbf{M})$.*

This allows us in particular to define for all $Z \in \mathrm{Irr}(\Pi(\mathbf{M}))^{\max}$ and $i \in I$ the value $\varphi_i(Z)$, see (4.1), such that for a dense open subset $U \subset Z$ we have $\varphi_i(M) = \varphi_i(Z)$ for all $M \in U$. Similarly, we can define $\varphi_i^*(Z)$.

Next we set

$$\mathrm{Irr}(\Pi(\mathbf{r})^{i,p})^{\max} := \{Z \in \mathrm{Irr}(\Pi(\mathbf{r})^{i,p}) \mid \dim Z = \dim H(\mathbf{r})\}$$

for $i \in I$ and $p \in \mathbb{N}_0$, and similarly $\mathrm{Irr}(\Pi(\mathbf{r})_{i,p}$. By Lemma 4.3 we get a bijection

$$e_i^*(\mathbf{r}, p)\colon \ \mathrm{Irr}(\Pi(\mathbf{r})^{i,p})^{\max} \to \mathrm{Irr}(\Pi(\mathbf{r} + \alpha_i)^{i,p+1})^{\max}, Z \mapsto p''(p'^{-1}(Z \times J_0))$$

Similarly we obtain a bijection

$$e_i(\mathbf{r}, p)\colon \ \mathrm{Irr}(\Pi(\mathbf{r})_{i,p})^{\max} \to \mathrm{Irr}(\Pi(\mathbf{r} + \alpha_i)_{i,p+1})^{\max}.$$

This allows us to define for all $\mathbf{r} \in \mathbb{N}^I$ the operators

$$\tilde{e}_i\colon \ \mathrm{Irr}(\Pi(\mathbf{r}))^{\max} \to \mathrm{Irr}(\Pi(\mathbf{r} + \alpha_i)), Z \mapsto \overline{e_i(\mathbf{r}, \varphi_i(Z))(Z^\circ)},$$

where $Z^\circ \in \mathrm{Irr}(\Pi(\mathbf{r})_{i,\varphi_i(Z)})^{\max}$ is the unique irreducible component with $\overline{Z^\circ} = Z$. Similarly, we can define the operators $\tilde{e}_i^*$ in terms of the bijections $e_i^*(\mathbf{r}, p)$. We define now

$$\mathfrak{B} := \coprod_{\mathbf{r} \in \mathbb{N}_0} \mathrm{Irr}(\Pi(\mathbf{r}))^{\max} \text{ and } \mathrm{wt}\colon \mathfrak{B} \to \mathbb{Z}^I, \ Z \mapsto \underline{\mathrm{rank}}(Z). \tag{4.2}$$

It is easy to see that $(\mathfrak{B}, \mathrm{wt}, (\tilde{e}_i, \varphi_i)_{i \in I})$ is special case of a lowest weight crystal in the sense of Kashiwara [1995, Sec. 7.2], namely we have

- $\varphi_i(\tilde{e}_i(b)) = \varphi_i(b) + 1, \ \mathrm{wt}(\tilde{e}_i(b)) = \mathrm{wt}(b) + \alpha_i$,
- with $\{b_-\} := \mathrm{Irr}(\Pi(0))^{\max}$, for each $b \in \mathcal{B}$ there exists a sequence $i_1, \dots, i_l$ of elements of I with $\tilde{e}_{i_1}\tilde{e}_{i_2}\cdots\tilde{e}_{i_l}(b_-) = b$,
- $\varphi_i(b) = 0$ implies $b \notin \mathrm{Im}(\tilde{e}_i)$.

Similarly $(\mathcal{B}, \mathrm{wt}, (\tilde{e}_i^*, \varphi_i^*)_{i\in I})$ is a lowest weight crystal with the same lowest weight element b_-.

Lemma 4.6. *The above defined operators and functions on $\mathcal{B}$ fulfill additionally the following conditions:*

(a) If $i \neq j$, then $\tilde{e}_i^*\tilde{e}_j(b) = \tilde{e}_j\tilde{e}_i^*(b)$.

(b) For all $b \in B$ we have $\varphi_i(b) + \varphi_i^*(b) - \langle \mathrm{wt}(b), \alpha_i\rangle \geq 0$.

(c) If $\varphi_i(b) + \varphi_i^*(b) - \langle \mathrm{wt}(b), \alpha_i\rangle = 0$, then $\tilde{e}_i(b) = \tilde{e}_i^*(b)$.

(d) If $\varphi_i(b) + \varphi_i^*(b) - \langle \mathrm{wt}(b), \alpha_i\rangle \geq 1$, then $\varphi_i(\tilde{e}_i^*(b)) = \varphi_i(b)$ and $\varphi_i^*(\tilde{e}_i(b)) = \varphi_i^*(b)$.

(e) If $\varphi_i(b) + \varphi_i^*(b) - \langle \mathrm{wt}(b), \alpha_i\rangle \geq 2$, then $\tilde{e}_i\tilde{e}_i^*(b) = \tilde{e}_i^*\tilde{e}_i(b)$.

The proof of this Lemma in Geiß, Leclerc, and Schröer [2017c, Sec. 5.6] uses the homological features of locally free Π-modules from Corollary 3.15 in an essential way. Note that here, by definition, $\langle \mathbf{r}, \alpha_i\rangle = (C\cdot\mathbf{r})_i$.

Altogether this means, by a criterion of Kashiwara and Saito [1997, Prop. 3.2.3], which we use here in a reformulation due to Tingley and Webster [2016, Prop. 1.4], that $(\mathcal{B}, \mathrm{wt}, (\tilde{e}_i, \varphi_i)_{i\in I}$ $(\mathcal{B}, \mathrm{wt}, (\tilde{e}_i^*, \varphi_i^*)_{i\in I}) \cong B_C(-\infty)$. Here, $B_C(-\infty)$ is the crystal graph of the quantum group $U_q(\mathfrak{n}(C))$. This implies part (b) of Theorem 4.2.

Remark 4.7. We did not give here Kashiwara's general definition of a crystal graph, or that of a lowest weight crystal associated to a dominant integral weight. The reason is that, due to limitations of space, we can not to set up the, somehow unwieldy, notations for the integral weights of a Kac-Moody Lie algebra. The interested reader can look up the relevant definitions, in a form which is compatible with these notes, in Geiß, Leclerc, and Schröer [2017c, Sec. 5.1, 5.2].

5 Algebras of constructible functions

5.1 Constructible functions and Euler characteristic. Recall that the topological Euler characteristic, defined in terms of singular cohomology with compact support and rational coefficients, defines a ring homomorphism from the Grothendieck ring of complex varieties to the integers.

By definition, a constructible function $f: X \to \mathbb{C}$ on a complex algebraic variety X has finite image, and $f^{-1}(c) \subset C$ is a constructible set for all $c \in \mathbb{C}$. By the above remark it makes sense to define

$$\int_{x \in X} f d\chi := \sum_{c \in \mathbb{C}} c\, \chi(f^{-1}(c)).$$

If $\varphi\colon X \to Y$ is a morphism of varieties, we can define the *push forward* of constructible functions via $(\varphi_*(f))(y) := \int_{x\in\varphi^{-1}(y)} f d\chi$. This is functorial in the sense that $(\psi \circ \varphi)_*(f) = \psi_*(\varphi_*(f))$ for $\psi\colon Y \to Z$ an other morphism, by result of MacPherson [1974, Prop. 1]. See also Joyce [2006, Sec. 3] for a careful discussion.

5.2 Convolution algebras as enveloping algebras. Let $A = \mathbb{C}Q$ as in Section 4.1. We consider for a dimension vector $\mathbf{d} \in \mathbb{N}^I$ the vector space $\mathfrak{F}(A)_{\mathbf{d}}$ of constructible functions $f\colon \operatorname{Rep}(A, \mathbf{d}) \to \mathbb{C}$ which are constant on $\mathrm{GL}_{\mathbf{d}}(\mathbb{C})$-orbits and set

$$\mathfrak{F}(A) := \bigoplus_{\mathbf{d}\in\mathbb{N}^i} \mathfrak{F}(A)_{\mathbf{d}}.$$

Following Lusztig [1991] $\mathfrak{F}(A)$ has the structure of a unitary, graded associative algebra. The multiplication is defined by

$$(f * g)(X) = \int_{U \in \mathrm{Gr}_{\mathbf{d}}^A(X)} f(U) g(X/U) d\chi,$$

where $f \in \mathfrak{F}(A)_{\mathbf{d}}$, $g \in \mathfrak{F}(A)_{\mathbf{e}}$, $X \in \operatorname{Rep}(A, \mathbf{d}+\mathbf{e})$, and $\mathrm{Gr}_{\mathbf{d}}^A(X)$ denotes the quiver Grassmannian of $\mathbf{d}$-dimensional subrepresentations of X. The associativity of the multiplication follows easily from the functoriality of the push-forward of constructible functions. We have an algebra homomorphism

$$c\colon \mathfrak{F}(A) \to \mathfrak{F}(A \times A), \text{ with } (c(f))(X, Y) = f(X \oplus Y), \tag{5.1}$$

see for example Geiß, Leclerc, and Schröer [2016, Sec. 4.3]. The proof depends crucially on the Białynicki-Birula result about the fixpoints of algebraic torus actions Białynicki-Birula [1973, Cor. 2]. This fails for example over the real numbers.

Remark 5.1. If $\mathbf{X} = (X_j)_{j\in J}$ is a family of indecomposable representations of A, we define the characteristic functions $\theta_j \in \mathfrak{F}_{\underline{\dim} X_j}(A)$ of the $\mathrm{GL}_{\underline{\dim} X_j}$-orbit $\mathcal{O}(X_j) \subset \operatorname{Rep}(A, \underline{\dim} X_j)$ and consider the graded subalgebra $\mathfrak{M}(A) = \mathfrak{M}_{\mathbf{X}}(A)$ of $\mathfrak{F}(A)$, which is generated by the θ_j. Clearly, the homogeneous components of $\mathfrak{M}$ are finite dimensional. If $\mathbf{j}$ =

$(j_1, j_2, \ldots, j_l)$ is a sequence of elements of j we have by the definition of the multiplication

$$\theta_{j_1} * \theta_{j_l} * \cdots * \theta_{j_l}(X) = \chi(\mathrm{Fl}^A_{\mathbf{X},\mathbf{j}}(M)),$$

where $\mathrm{Fl}^A_{\mathbf{X},\mathbf{j}}(M)$ denotes the variety of all flags of submodules

$$0 = M^{(0)} \subset M^{(1)} \subset \cdots \subset M^{(l)} = M$$

with $M^{(k)}/M^{(k-1)} \cong X_{j_k}$ for $k = 1, 2, \ldots, l$. In particular, if M has no filtration with all factors isomorphic to some X_j, we have $f(M) = 0$ for all $f \in \mathfrak{M}(A)_{\underline{\dim} M}$. See Geiß, Leclerc, and Schröer [2016, Lemma 4.2].

Lemma 5.2. *The morphism c from (5.1) induces a comultiplication $\Delta\colon \mathfrak{M}(A) \to \mathfrak{M}(A) \otimes \mathfrak{M}(A)$ with $\Delta(\theta_j) = \theta_j \otimes 1 + 1 \otimes \theta_j$ for all j. With this structure $\mathfrak{M}$ is a cocommutative Hopf algebra, which is isomorphic to the enveloping algebra $U(\mathcal{P}(\mathfrak{M}))$ of the Lie algebra of its primitive elements $\mathcal{P}(\mathfrak{M})$.*

See Geiß, Leclerc, and Schröer [ibid., Prop. 4.5] for a proof. Recall, that an element x of a Hopf algebra is called *primitive* iff $\Delta(x) = x \otimes 1 + 1 \otimes x$. It is straightforward to check that the primitive elements of a Hopf algebra form a Lie algebra under the usual commutator $[x, y] = xy - yx$.

Remark 5.3. It is important to observe that, by the definition of the comultiplication, the support of any primitive element of $\mathfrak{M}$ consists of indecomposable, $\mathbf{X}$-filtered modules. In fact, for $f \in \mathcal{P}(\mathfrak{M})$ and $M, N \in \mathrm{rep}(A)$ we have $f(M \oplus N) = cf(M, N) = (f \otimes 1 + 1 \otimes f)(M, N)$. See Geiß, Leclerc, and Schröer [ibid., Lem. 4.6].

We are here interested in the two special cases when $A = H_{\mathbb{C}}(C, D, \Omega)$ or $A = \Pi_{\mathbb{C}}(C, D)$ and $\mathbf{X} = \mathbf{E} = (E_i)_{i \in I}$. Note that by Remark 5.1 only locally free modules can appear in the support of any $f \in \mathfrak{M}_{\mathbf{E}}(H)$. Similarly, the support of any $f \in \mathfrak{M}_{\mathbf{E}}(\Pi)$ consists only of $\mathbf{E}$-filtered modules. For this reason we will consider in what follows, both $\mathfrak{M}_{\mathbf{E}}(H)$ and $\mathfrak{M}_{\mathbf{E}}(\Pi)$ as graded by *rank vectors*. In other words, from now on

$$\mathfrak{M}(H) := \mathfrak{M}_{\mathbf{E}}(H) = \bigoplus_{\mathbf{r} \in \mathbb{N}^I} \mathfrak{M}_{\mathbf{r}}(H) \quad \text{and} \quad \mathfrak{M}(\Pi) = \mathfrak{M}_{\mathbf{E}}(\Pi) = \bigoplus_{\mathbf{r} \in \mathbb{N}^I} \mathfrak{M}_{\mathbf{r}}(\Pi),$$

where we may consider the the elements of $\mathfrak{M}_{\mathbf{r}}(H) := \mathfrak{M}_{\mathbf{E}}(H)_{D \cdot \mathbf{r}}$ as constructible functions on $H(\mathbf{r})$. Similarly we may consider the elements of $\mathfrak{M}_{\mathbf{r}}(\Pi) := \mathfrak{M}_{\mathbf{E}}(\Pi)_{D \cdot \mathbf{r}}$ as constructible functions on $\Pi(\mathbf{r})$.

5.3 $\mathfrak{M}_{\mathbf{E}}(H)$ and a dual PBW-basis in the Dynkin case. We have the following basic result from Geiß, Leclerc, and Schröer [ibid., Cor. 4.10].

Proposition 5.4. *Let C be a symmetrizable Cartan matrix, D a symmetrizer and Ω an orientation for C. With $H = H_{\mathbb{C}}(C, D, \Omega)$ we have an surjective Hopf algebra homomorphism*

$$\eta_H \colon U(\mathfrak{n}(C)) \to \mathcal{M}_{\mathbf{E}}(H) \text{ defined by } \quad e_i \mapsto \theta_i (i \in I).$$

The main point is to show that for the θ_i $(i \in I)$ fulfill the Serre relations (2.4). More precisely we need that the primitive elements

$$\theta_{ij} := (\operatorname{ad} \theta_i)^{1-c_{ij}}(\theta_j) \in \mathcal{P}(\mathcal{M}(H))_{(1-c_{ij})\alpha_i+\alpha_j} \quad (i \neq j)$$

actually vanish. For this it is enough, by Remark 5.3, to show that there exists no *indecomposable*, locally free H-module M with $\underline{\operatorname{rank}}(M) = (1-c_{ij})\alpha_i + \alpha_j$. This is carried out in the proof of Geiß, Leclerc, and Schröer [2016, Prop. 4.9].

The proof of the following result, which is Geiß, Leclerc, and Schröer [ibid., Thm. 6.1], occupies the major part of that paper.

Proposition 5.5. *Let C be a symmetrizable Cartan matrix of Dynkin type, D a symmetrizer and Ω an orientation for C and $H = H_{\mathbb{C}}(C, D, \Omega)$. Then for each positive root $\beta \in \Delta^+$ there exists a primitive element $\theta_\beta \in \mathcal{P}(\mathcal{M}(H))_\beta$ with $\theta_\beta(M(\beta)) = 1$.*

The idea of the proof is as follows: By Geiß, Leclerc, and Schröer [2017b, Cor. 1.3] for any $\beta \in \Delta^+(C)$ and any sequence $\mathbf{i}$ in I, the Euler characteristic $\chi(\mathrm{Fl}^H_{\mathbf{E},\mathbf{i}}(M(\beta)))$ is independent of the choice of the symmetrizer D. So, we may assume that C is connected and D minimal. In the symmetric (quiver) case, our claim follows now by Schofield's result Schofield [n.d.], who showed that in this case $\mathcal{P}(\mathcal{M}(H))$ can be identified with $\mathfrak{n}(C)$. By Gabriel's theorem in this case the θ_β are the characteristic function of the GL_β-orbit of $M(\beta)$.

In the remaining cases, we construct the θ_β by induction on the height of β in terms of (iterated) commutators of "smaller" θ_γ. Note however that in this case this construction is delicate since the support of the θ_β may contain several indecomposable, locally free modules. See for example Geiß, Leclerc, and Schröer [2017d, Sec. 13.2(d)].

Since in the Dynkin case all weight spaces of $\mathfrak{n}(C)$ are one-dimensional, the main result of Geiß, Leclerc, and Schröer [2016], Theorem 1.1 (ii), follows easily:

Theorem 5.6. *If C is of Dynkin type, the Hopf algebra homomorphism η_H is an isomorphism.*

Recall the notation used in Proposition 3.11. In particular, $\mathbf{i}$ is a reduced expression for the longest element $w_0 \in W(C)$, which is $+$-adapted to Ω, and $\beta_k = \beta_{\mathbf{i},k}$ for $k = 1, 2, \ldots, r$. Let us abbreviate

$$\theta_{\mathbf{m}} := \frac{1}{m_r! \cdots m_1!} \theta_{\beta_r}^{m_r} * \cdots * \theta_{\beta_1}^{m_1} \quad \text{and} \quad M(\mathbf{m}) := \oplus_{k=1}^r M(\beta_k)^{m_k}$$

for $\mathbf{m} = (m_1, m_2, \dots, m_r) \in \mathbb{N}^r$. By the above results $(\theta_{\mathbf{m}})_{\mathbf{m}\in\mathbb{N}^r}$ is a normalized PBW-basis of $\mathfrak{M}(H) \cong U(\mathfrak{n}(C))$ in the Dynkin case.

Moreover we consider the graded dual $\mathfrak{M}(H)^*$ of $\mathfrak{M}(H)$, and the evaluation form $\delta_{M(\mathbf{m})} \in \mathfrak{M}(H)^*$ with $\delta_{\mathbf{m}}(f) := f(M(\mathbf{m}))$. By the definition of the comultiplication in $\mathfrak{M}(H)$, the graded dual is a commutative Hopf algebra, and $\delta_{M(\mathbf{m})} \cdot \delta_{M(\mathbf{n})} = \delta_{M(\mathbf{m}+\mathbf{n})}$. Our next result is essentially Geiß, Leclerc, and Schröer [2017d, Thm. 1.3].

Proposition 5.7. *With the above notation we have*

$$\delta_{M(\mathbf{m})}(\theta_{\mathbf{n}}) = \delta_{\mathbf{m},\mathbf{n}} \textit{ for all } \mathbf{m}, \mathbf{n} \in \mathbb{N}^r.$$

Thus $(\delta_{M(\mathbf{m})})_{\mathbf{m}\in\mathbb{N}^r}$ is a basis of $\mathfrak{M}(H)^*$ which is dual to the PBW-basis $(\theta_{\mathbf{m}})_{\mathbf{m}\in\mathbb{N}^r}$, and $\mathfrak{M}(H)^* = \mathbb{C}[\delta_{M(\beta_1)}, \dots, \delta_{M(\beta_r)}]$.

In the quiver case (with trivial symmetrizer) this result is easy to prove, since with Gabriel's theorem and Proposition 3.11 follows quickly that $\theta_{\mathbf{r}}$ is the characteristic function of the orbit of $M(\mathbf{r})$. However, in our more general setting, already the θ_{β_k} are usually not the characteristic function of $M(\beta_k)$, as we observed above. The more sophisticated Proposition 3.12 implies, by the definition of the multiplication in $\mathfrak{M}_{\mathbf{E}}(H)$, that $\theta_{\mathbf{m}}(M(\beta_k)) = 0$ if $\mathbf{m} \neq \mathbf{e}_k$, the k-th unit vector. The remaining claims follow now by formal arguments, see the proof of Geiß, Leclerc, and Schröer [ibid., Thm. 6.1].

Remark 5.8. For $M \in \operatorname{rep}_{\mathrm{l.f.}}(H)$ and $\mathbf{e} \in \mathbb{N}^I$ we we introduce the quasi-projective variety

$$\operatorname{Grlf}^H_{\mathbf{e}}(M) := \{U \subset M \mid U \text{ locally free submodule and } \underline{\operatorname{rank}}(U) = \mathbf{e}\},$$

which is an open subset of the usual quiver Grassmannian $\operatorname{Gr}^H_{D\cdot\mathbf{e}}(M)$. With this notation we can define

$$F_M := \sum_{\mathbf{e}\in\mathbb{N}^I} \chi(\operatorname{Grlf}^H_{\mathbf{e}}(M)) Y^{\mathbf{e}} \in \mathbb{Z}[Y_1, \dots, Y_n] \text{ and } \mathbf{g}_M := -R \cdot \underline{\operatorname{rank}}(M),$$

where R is the matrix introduced in Corollary 3.4. By the main result of Geiß, Leclerc, and Schröer [ibid.] this yields for $M = M(\beta)$ with $\beta \in \Delta^+(C)$ the F-polynomial and g-vector, in the sense of Fomin and Zelevinsky [2007], for all cluster variables of a finite type cluster algebra Fomin and Zelevinsky [2003] of type C with respect to an acyclic seed defined by Ω. The proof is based on Proposition 5.7, and on the description by Yang and Zelevinsky [2008] of the F-polynomial of a cluster variable in terms of generalized minors.

5.4 Semicanonical functions and the support conjecture for $\mathcal{M}_{\mathbf{E}}(\Pi)$. Recall, that we abbreviate $\Pi = \Pi_{\mathbb{C}}(C, D)$ for a symmetrizable Cartan matrix C with symmetrizer D. By definition $\mathcal{M}(\Pi) = \mathcal{M}_{\mathbf{E}}(\Pi) \subset \mathcal{F}(\Pi)$ is generated by the functions $\tilde{\theta}_i \in \mathcal{M}_{\alpha_i}(\Pi)$ for $i \in I$, where $\tilde{\theta}_i$ is the characteristic function of the orbit of E_i, viewed as a Π-module. We use here the notation $\tilde{\theta}_i$ rather than θ_i to remind us that the multiplication is now defined in terms of constructible functions on a larger space. More precisely, we have for each $\mathbf{r} \in \mathbb{N}^I$ an injective $\mathrm{Aut}_S(\mathbf{E}^{\mathbf{r}})$-equivariant, injective morphism of varieties

$$\iota_{\mathbf{r}}\colon H(\mathbf{r}) \to \Pi(\mathbf{d}).$$

These morphisms induce, via restriction, a surjective morphism of graded Hopf algebras

$$\iota_{\Omega}^*\colon \mathcal{M}(\Pi) \to \mathcal{M}(H), \quad \tilde{\theta}_i \mapsto \theta_i \text{ for } i \in I.$$

The proof of the following result is, almost verbatim, the same induction argument as the one used by Lusztig [2000], see Geiß, Leclerc, and Schröer [2017c, Lem. 7.1].

Lemma 5.9. *Let $\mathbf{r} \in \mathbb{N}^I$. For each $Z \in \mathrm{Irr}(\Pi(\mathbf{r}))^{\max}$ there exists an open dense subset $U_Z \subset Z$ and a function $f_Z \in \mathcal{M}_{\mathbf{r}}(\Pi)$ such that for $Z, Z' \in \mathrm{Irr}(\Pi(\mathbf{r}))^{\max}$ and any $u' \in U_{Z'}$ we have*

$$f_Z(u') = \delta_{Z,Z'}.$$

In particular, the functions $(f_Z)_{Z\in\mathrm{Irr}(\Pi(\mathbf{r}))^{\max}}$ are linearly independent in $\mathcal{M}_{\mathbf{r}}(\Pi)$.

Note however, that the result is not trivial since we claim that the $f_Z \in \mathcal{M}_{\mathbf{e}}(\Pi)$ and not in the much bigger space $\mathcal{F}(\Pi)_{C\cdot\mathbf{r}}$. On the other hand, it is important to observe that the inductive construction of the *semicanonical functions* f_Z involves some choices.

As in Section 5.3, we define now for each $i \neq j$ in I the primitive element

$$\tilde{\theta}_{ij} = (\mathrm{ad}\,\tilde{\theta}_i)^{1-c_{ij}}(\tilde{\theta}_j) \in \mathcal{P}(\mathcal{M}(\Pi)).$$

Unfortunately, we have the following result, which is a combination of Lemma 6.1, Proposition 6.2 and Lemma 6.3 from Geiß, Leclerc, and Schröer [ibid.].

Lemma 5.10. *Suppose with the above notations that $c_{ij} < 0$.*

(a) If $c_i \geq 2$ then there exists an indecomposable, $\Pi = \Pi(C, D)$-module $X = X_{(ij)}$ with $\underline{\mathrm{rank}}(X_{(ij)}) = (1 - c_{ij})\alpha_i + \alpha_j$ and $\tilde{\theta}_{ij}(X_{(ij)}) \neq 0$.

(b) If M is crystal module with $\underline{\mathrm{rank}}(M) = (1 - c_{ij})\alpha_i + \alpha_j$ we have $\tilde{\theta}_{ij}(M) = 0$.

This leads us to define in $\mathcal{M}(\Pi)$ the ideal $\mathcal{J}$, which is generated by the homogeneous elements $\tilde{\theta}_{ij}$ for $i, j \in I$ with $i \neq j$. We set moreover

$$\overline{\mathcal{M}}(\Pi) = \mathcal{M}(\Pi)/\mathcal{J} \text{ and } \bar{f} := f + \mathcal{J} \quad (f \in \mathcal{M}(\Pi)).$$

Thus, by Proposition 5.4, the morphism ι_Ω^* induces a surjective algebra homomorphism $\overline{\iota}_\Omega^*\colon \overline{\mathfrak{M}}(\Pi) \to \mathfrak{M}(H)$. On the other hand, we can define for each $\mathbf{r} \in \mathbb{N}^I$ the space of functions with non maximal support

$$\mathcal{S}_\mathbf{r} := \{f \in \mathfrak{M}_\mathbf{r}(\Pi) \mid \dim \operatorname{supp}(f) < \dim H(\mathbf{r})\} \text{ and } \mathcal{S} := \oplus_{\mathbf{r}\in\mathbb{N}^I} \mathcal{S}_\mathbf{r}.$$

Recall that $\dim \Pi(\mathbf{r}) = \dim H(\mathbf{r})$. Proposition 4.5 and Lemma 5.10 imply at least that $\tilde{\theta}_{ij} \in \mathcal{S}$. In view of Lemma 5.9 and Theorem 4.2 it is easy to show the following result:

Proposition 5.11. *The following three conditions are equivalent:*

$$(1)\ \mathcal{I} \subset \mathcal{S}, \qquad (2)\ \mathcal{I} = \mathcal{S}, \qquad (3)\ \mathcal{S} \text{ is an ideal.}$$

In this case the surjective algebra homomorphism

$$\eta\colon U(\mathfrak{n}) \to \overline{\mathfrak{M}}(\Pi), e_i \mapsto \tilde{\theta}_i + \mathcal{I}$$

would be an isomorphism, and the $(\eta^{-1}(\bar{f}_Z))_{\mathfrak{B}}$ would form a basis of $U(\mathfrak{n})$ which is independent of the possible choices for the $(f_Z)_{Z\in\mathfrak{B}}$.

Thus we call the equivalent conditions of the above proposition our *Support conjecture*.

Remark 5.12. Our semicanonical basis would yield, similarly to Lusztig [2000, Sec. 3], in a natural way a basis for each integrable highest weight representation $L(\lambda)$ of $\mathfrak{g}(C)$, if the support conjecture is true. See Geiß, Leclerc, and Schröer [2017c, Se. 7.3] for more details.

Acknowledgments. I wish to thank Bernard Leclerc and Jan Schröer for their very precious collaboration and friendship over the years. I also should like to acknowledge hospitality from the MPIM-Bonn, where this manuscript was prepared.

References

Maurice Auslander, María Inés Platzeck, and Idun Reiten (1979). "Coxeter functors without diagrams". *Trans. Amer. Math. Soc.* 250, pp. 1–46. MR: 530043 (cit. on p. 125).

A. Białynicki-Birula (1973). "On fixed point schemes of actions of multiplicative and additive groups". *Topology* 12, pp. 99–103. MR: 0313261 (cit. on p. 136).

Sheila Brenner, Michael C. R. Butler, and Alastair D. King (2002). "Periodic algebras which are almost Koszul". *Algebr. Represent. Theory* 5.4, pp. 331–367. MR: 1930968 (cit. on p. 128).

Tom Bridgeland and Valerio Toledano Laredo (2012). "Stability conditions and Stokes factors". *Invent. Math.* 187.1, pp. 61–98. MR: 2874935.

Sergio Cecotti and Michele Del Zotto (2012). “4d $\mathcal{N} = 2$ gauge theories and quivers: the non-simply laced case”. *J. High Energy Phys.* 10, 190p. MR: 3033809 (cit. on p. 118).

William Crawley-Boevey (2000). “On the exceptional fibres of Kleinian singularities”. *Amer. J. Math.* 122.5, pp. 1027–1037. MR: 1781930 (cit. on p. 129).

William Crawley-Boevey and Peter Shaw (2006). “Multiplicative preprojective algebras, middle convolution and the Deligne-Simpson problem”. *Adv. Math.* 201.1, pp. 180–208. MR: 2204754 (cit. on p. 128).

Vlastimil Dlab and Claus Michael Ringel (1974). *Representations of graphs and algebras*. Carleton Mathematical Lecture Notes, No. 8. Department of Mathematics, Carleton University, Ottawa, Ont., pp. iii+86. MR: 0387350 (cit. on pp. 118, 123).

– (1979). “A module theoretical interpretation of properties of the root system”. In: *Ring theory (Proc. Antwerp Conf. (NATO Adv. Study Inst.), Univ. Antwerp, Antwerp, 1978)*. Vol. 51. Lecture Notes in Pure and Appl. Math. Dekker, New York, pp. 435–451. MR: 563306 (cit. on p. 127).

– (1980). “The preprojective algebra of a modulated graph”. In: *Representation theory, II (Proc. Second Internat. Conf., Carleton Univ., Ottawa, Ont., 1979)*. Vol. 832. Lecture Notes in Math. Springer, Berlin-New York, pp. 216–231. MR: 607155 (cit. on p. 118).

Sergey Fomin (2010). “Total positivity and cluster algebras”. In: *Proceedings of the International Congress of Mathematicians. Volume II*. Hindustan Book Agency, New Delhi, pp. 125–145. MR: 2827788 (cit. on p. 117).

Sergey Fomin and Andrei Zelevinsky (2003). “Cluster algebras. II. Finite type classification”. *Invent. Math.* 154.1, pp. 63–121. MR: 2004457 (cit. on p. 139).

– (2007). “Cluster algebras. IV. Coefficients”. *Compos. Math.* 143.1, pp. 112–164. MR: 2295199 (cit. on p. 139).

Peter Gabriel (1980). “Auslander-Reiten sequences and representation-finite algebras”. In: *Representation theory, I (Proc. Workshop, Carleton Univ., Ottawa, Ont., 1979)*. Vol. 831. Lecture Notes in Math. Springer, Berlin, pp. 1–71. MR: 607140 (cit. on p. 126).

Christof Geiß, Bernard Leclerc, and Jan Schröer (2011). “Kac-Moody groups and cluster algebras”. *Adv. Math.* 228.1, pp. 329–433. MR: 2822235 (cit. on p. 117).

– (2012). “Generic bases for cluster algebras and the Chamber ansatz”. *J. Amer. Math. Soc.* 25.1, pp. 21–76. MR: 2833478 (cit. on p. 117).

– (2016). “Quivers with relations for symmetrizable Cartan matrices III: Convolution algebras”. *Represent. Theory* 20, pp. 375–413. MR: 3555157 (cit. on pp. 136–138).

– (2017a). “Quivers with relations for symmetrizable Cartan matrices I: Foundations”. *Invent. Math.* 209.1, pp. 61–158. MR: 3660306 (cit. on pp. 117, 118, 123–126, 128, 129).

– (2017b). “Quivers with relations for symmetrizable Cartan matrices II: Change of symmetrizer”. *Int. Math. Res. Not.* 33p. (Cit. on p. 138).

– (Feb. 2017c). "Quivers with relations for symmetrizable Cartan matrices IV: Crystal graphs and semicanonical functions". To appear in Selecta Math. arXiv: 1702.07570 (cit. on pp. 121, 131, 132, 134, 135, 140, 141).

– (Apr. 2017d). "Quivers with relations for symmetrizable Cartan matrices V: Caldero-Chapoton formula". To appear in Proc. London Math. Soc. arXiv: 1704.06438 (cit. on pp. 117, 127, 138, 139).

I. M. Gelfand and V. A. Ponomarev (1979). "Model algebras and representations of graphs". *Funktsional. Anal. i Prilozhen.* 13.3, pp. 1–12. MR: 545362 (cit. on p. 118).

J. Geuenich (2016). "Quiver Modulations and Potentials". PhD thesis. Bonn, p. 221 (cit. on p. 118).

David Hernandez and Bernard Leclerc (2016). "A cluster algebra approach to q-characters of Kirillov-Reshetikhin modules". *J. Eur. Math. Soc. (JEMS)* 18.5, pp. 1113–1159. arXiv: 1303.0744. MR: 3500832 (cit. on p. 117).

Dominic Joyce (2006). "Constructible functions on Artin stacks". *J. London Math. Soc. (2)* 74.3, pp. 583–606. MR: 2286434 (cit. on p. 136).

Victor G. Kac (1990). *Infinite-dimensional Lie algebras*. Third. Cambridge University Press, Cambridge, pp. xxii+400. MR: 1104219 (cit. on p. 121).

M. Kashiwara (1991). "On crystal bases of the Q-analogue of universal enveloping algebras". *Duke Math. J.* 63.2, pp. 465–516. MR: 1115118.

Masaki Kashiwara (1995). "On crystal bases". In: *Representations of groups (Banff, AB, 1994)*. Vol. 16. CMS Conf. Proc. Amer. Math. Soc., Providence, RI, pp. 155–197. MR: 1357199 (cit. on p. 134).

Masaki Kashiwara and Yoshihisa Saito (1997). "Geometric construction of crystal bases". *Duke Math. J.* 89.1, pp. 9–36. MR: 1458969 (cit. on p. 135).

G. Lusztig (1991). "Quivers, perverse sheaves, and quantized enveloping algebras". *J. Amer. Math. Soc.* 4.2, pp. 365–421. MR: 1088333 (cit. on pp. 132, 136).

– (2000). "Semicanonical bases arising from enveloping algebras". *Adv. Math.* 151.2, pp. 129–139. MR: 1758244 (cit. on pp. 140, 141).

R. D. MacPherson (1974). "Chern classes for singular algebraic varieties". *Ann. of Math. (2)* 100, pp. 423–432. MR: 0361141 (cit. on p. 136).

Vinoth Nandakumar and Peter Tingley (2016). "Quiver varieties and crystals in symmetrizable type via modulated graphs". *Math. Res. Lett. (to appear)*. arXiv: 1606.01876 (cit. on p. 118).

S. Omlor (2016). "A module theoretical interpretation of properties of the root systems". Bonn. MA thesis (cit. on p. 127).

M Popa (2011). "Modern aspects of the cohomological study of varieties". *Lecture notes* 3, p. 20.

Claus Michael Ringel (1984). *Tame algebras and integral quadratic forms*. Vol. 1099. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. xiii+376. MR: 774589 (cit. on p. 118).

– (1998). "The preprojective algebra of a quiver". In: *Algebras and modules, II (Geiranger, 1996)*. Vol. 24. CMS Conf. Proc. Amer. Math. Soc., Providence, RI, pp. 467–480. MR: 1648647.

A. Schofield (n.d.). "Quivers and Kac-Moody Lie algebras" (cit. on p. 138).

Peter Tingley and Ben Webster (2016). "Mirković-Vilonen polytopes and Khovanov-Lauda-Rouquier algebras". *Compos. Math.* 152.8, pp. 1648–1696. MR: 3542489 (cit. on p. 135).

Daisuke Yamakawa (2010). "Quiver varieties with multiplicities, Weyl groups of non-symmetric Kac-Moody algebras, and Painlevé equations". *SIGMA Symmetry Integrability Geom. Methods Appl.* 6, Paper 087, 43pp. MR: 2769928 (cit. on p. 118).

Shih-Wei Yang and Andrei Zelevinsky (2008). "Cluster algebras of finite type via Coxeter elements and principal minors". *Transform. Groups* 13.3-4, pp. 855–895. MR: 2452619 (cit. on p. 139).

Received 2017-12-12.

Christof Geiß
Instituto de Matemáticas
Universidad Nacional Autónoma de México
Ciudad Universitaria
04510 Cd. de México
Mexico
christof@im.unam.mx
christof.geiss@im.unam.mx

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (143–180)

TILTING COHEN–MACAULAY REPRESENTATIONS

OSAMU IYAMA

Dedicated to the memory of Ragnar-Olaf Buchweitz

Abstract

This is a survey on recent developments in Cohen-Macaulay representations via tilting and cluster tilting theory. We explain triangle equivalences between the singularity categories of Gorenstein rings and the derived (or cluster) categories of finite dimensional algebras.

1 Introduction

The study of Cohen-Macaulay (CM) representations (Curtis and Reiner [1981], Yoshino [1990], Simson [1992], and Leuschke and Wiegand [2012]) is one of the active subjects in representation theory and commutative algebra. It has fruitful connections to singularity theory, algebraic geometry and physics. This article is a survey on recent developments in this subject.

The first half of this article is spent for background materials, which were never written in one place. In Section 2, we recall the notion of CM modules over Gorenstein rings, and put them into the standard framework of triangulated categories. This gives us powerful tools including Buchweitz's equivalence between the stable category $\underline{\mathrm{CM}}R$ and the singularity category, and Orlov's realization of the graded singularity category in the derived category, giving a surprising connection between CM modules and algebraic geometry. We also explain basic results including Auslander-Reiten duality stating that $\underline{\mathrm{CM}}R$ is a Calabi-Yau triangulated category for a Gorenstein isolated singularity R, and Gabriel's Theorem on quiver representations and its commutative counterpart due to Buchweitz-Greuel-Schreyer.

In Section 3, we give a brief introduction to tilting and cluster tilting. Tilting theory controls equivalences of derived categories, and played a central role in Cohen-Macaulay

The author was partially supported by JSPS Grant-in-Aid for Scientific Research (B) 16H03923 and (S) 15H05738.
MSC2010: primary 16G50; secondary 13C14, 16E35, 18E30, 14F05.

approximation theory around 1990 (Auslander and Buchweitz [1989] and Auslander and Reiten [1991]). The first main problem of this article is to find a tilting object in the stable category $\underline{\mathrm{CM}}^G R$ of a G-graded Gorenstein ring R. This is equivalent to find a triangle equivalence

$$\underline{\mathrm{CM}}^G R \simeq \mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\Lambda) \tag{1-1}$$

with some ring Λ. It reveals a deep connection between rings R and Λ.

The notion of d-cluster tilting was introduced in higher Auslander-Reiten theory. A Gorenstein ring R is called d-CM-finite if there exists a d-cluster tilting object in CM R. This property is a natural generalization of CM-finiteness, and closely related to the existence of non-commutative crepant resolutions of Van den Bergh. On the other hand, the d-cluster category $\mathrm{C}_d(\Lambda)$ of a finite dimensional algebra Λ is a d-Calabi-Yau triangulated category containing a d-cluster tilting object, introduced in categorification of Fomin-Zelevinsky cluster algebras. The second main problem of this article is to find a triangle equivalence

$$\underline{\mathrm{CM}} R \simeq \mathrm{C}_d(\Lambda) \tag{1-2}$$

with some finite dimensional algebra Λ. This implies that R is d-CM-finite.

In the latter half of this article, we construct various triangle equivalences of the form (1-1) or (1-2). In Section 4, we explain results in Yamaura [2013] and Buchweitz, Iyama, and Yamaura [2018]. They assert that, for a large class of $\mathbb{Z}$-graded Gorenstein rings R in dimension 0 or 1, there exist triangle equivalences (1-1) for some algebras Λ.

There are no such general results in dimension greater than 1. Therefore in the main Sections 5 and 6 of this article, we concentrate on special classes of Gorenstein rings. In Section 5, we explain results on Gorenstein rings obtained from classical and higher preprojective algebras (Amiot, Iyama, and Reiten [2015], Iyama and Oppermann [2013], and Kimura [2018, 2016]). A crucial observation is that certain Calabi-Yau algebras are higher preprojective algebras and higher Auslander algebras at the same time. In Section 6, we explain results on CM modules on Geigle-Lenzing complete intersections and the derived categories of coherent sheaves on the associated stacks (Herschend, Iyama, Minamoto, and Oppermann [2014]). They are higher dimensional generalizations of weighted projective lines of Geigle-Lenzing.

Acknowledgements The author would like to thank Claire Amiot, Bernhard Keller and Idun Reiten for valuable comments on this article.

2 Preliminaries

2.1 Notations. We fix some conventions in this paper. All modules are right modules. The composition of $f: X \to Y$ and $g: Y \to Z$ is denoted by gf. For a ring Λ, we denote by $\operatorname{mod}\Lambda$ the category of finitely generated Λ-modules, by $\operatorname{proj}\Lambda$ the category of finitely generated projective Λ-modules, and by $\operatorname{gl.dim}\Lambda$ the global dimension of Λ. When Λ is G-graded, we denote by $\operatorname{mod}^G\Lambda$ and $\operatorname{proj}^G\Lambda$ the G-graded version, whose morphisms are degree preserving. We denote by k an arbitrary field unless otherwise specified, and by D the k-dual or Matlis dual over a base commutative ring.

2.2 Cohen-Macaulay modules. We start with the classical notion of Cohen-Macaulay modules over commutative rings (Bruns and Herzog [1993] and Matsumura [1989]).

Let R be a commutative noetherian ring. The *dimension* $\dim R$ of R is the supremum of integers $n \geq 0$ such that there exists a chain $\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n$ of prime ideals of R. The *dimension* $\dim M$ of $M \in \operatorname{mod} R$ is the dimension $\dim(R/\operatorname{ann} M)$ of the factor ring $R/\operatorname{ann} M$, where $\operatorname{ann} M$ is the annihilator of M.

The notion of depth is defined locally. Assume that R is a local ring with maximal ideal $\mathfrak{m}$ and $M \in \operatorname{mod} R$ is non-zero. An element $r \in \mathfrak{m}$ is called *M-regular* if the multiplication map $r: M \to M$ is injective. A sequence $r_1, \ldots, r_n$ of elements in $\mathfrak{m}$ is called an *M-regular sequence* of length n if r_i is $(M/(r_1, \ldots, r_{i-1})M)$-regular for all $1 \leq i \leq n$. The *depth* $\operatorname{depth} M$ of M is the supremum of the length of M-regular sequences. This is given by the simple formula

$$\operatorname{depth} M = \inf\{i \geq 0 \mid \operatorname{Ext}^i_R(R/\mathfrak{m}, M) \neq 0\}.$$

The inequalities $\operatorname{depth} M \leq \dim M \leq \dim R$ hold. We call M (maximal) *Cohen-Macaulay* (or *CM*) if the equality $\operatorname{depth} M = \dim R$ holds or $M = 0$.

When R is not necessarily local, $M \in \operatorname{mod} R$ is called *CM* if $M_{\mathfrak{m}}$ is a CM $R_{\mathfrak{m}}$-module for all maximal ideals $\mathfrak{m}$ of R. The ring R is called *CM* if it is CM as an R-module. The ring R is called *Gorenstein* (resp. *regular*) if $R_{\mathfrak{m}}$ has finite injective dimension as an $R_{\mathfrak{m}}$-module (resp. $\operatorname{gl.dim} R_{\mathfrak{m}} < \infty$) for all maximal ideals $\mathfrak{m}$ of R. In this case, the injective (resp. global) dimension coincides with $\dim R_{\mathfrak{m}}$, but this is not true in the more general setting below. The following hierarchy is basic.

$$\text{Regular rings} \Longrightarrow \text{Gorenstein rings} \Longrightarrow \text{Cohen-Macaulay rings}$$

We will study CM modules over Gorenstein rings. Since we apply methods in representation theory, it is more reasonable to work in the following wider framework.

Definition 2.1 (Iwanaga [1979] and Enochs and Jenda [2000])**.** Let Λ be a (not necessarily commutative) noetherian ring, and $d \geq 0$ an integer. We call Λ (d-)*Iwanaga-Gorenstein* (or *Gorenstein*) if Λ has injective dimension at most d as a Λ-module, and also as a Λ^{op}-module.

Clearly, a commutative noetherian ring R is Iwanaga-Gorenstein if and only if it is Gorenstein and $\dim R < \infty$. Note that there are various definitions of non-commutative Gorenstein rings, e.g. Artin and Schelter [1987], Curtis and Reiner [1981], Fossum, Griffith, and Reiten [1975], Goto and Nishida [2002], and Iyama and Wemyss [2014]. Although Definition 2.1 is much weaker than them, it is sufficient for the aim of this paper.

Noetherian rings with finite global dimension are analogues of regular rings, and form special classes of Iwanaga-Gorenstein rings. The first class consists of *semisimple rings* (i.e. rings Λ with $\operatorname{gl.dim} \Lambda = 0$), which are products of matrix rings over division rings by Artin-Wedderburn Theorem. The next class consists of *hereditary rings* (i.e. rings Λ with $\operatorname{gl.dim} \Lambda \leq 1$), which are obtained from quivers.

Definition 2.2 (Assem, Simson, and Skowroński [2006])**.** A *quiver* is a quadruple $Q = (Q_0, Q_1, s, t)$ consisting of sets Q_0, Q_1 and maps $s, t \colon Q_1 \to Q_0$. We regard each element in Q_0 as a vertex, and $a \in Q_1$ as an arrow with source $s(a)$ and target $t(a)$. A *path* of length 0 is a vertex, and a *path* of length $\ell (\geq 1)$ is a sequence $a_1 a_2 \cdots a_\ell$ of arrows satisfying $t(a_i) = s(a_{i+1})$ for each $1 \leq i < \ell$.

For a field k, the *path algebra* kQ is defined as follows: It is a k-vector space with basis consisting of all paths on Q. For paths $p = a_1 \cdots a_\ell$ and $q = b_1 \cdots b_m$, we define $pq = a_1 \cdots a_\ell b_1 \cdots b_m$ if $t(a_\ell) = s(b_1)$, and $pq = 0$ otherwise.

Clearly $\dim_k(kQ)$ is finite if and only if Q is *acyclic* (that is, there are no paths p of positive length satisfying $s(p) = t(p)$).

Example 2.3. (a) (Assem, Simson, and Skowroński [ibid.]) The path algebra kQ of a finite quiver Q is hereditary. Conversely, any finite dimensional hereditary algebra over an algebraically closed field k is Morita equivalent to kQ for some acyclic quiver Q.

(b) A finite dimensional k-algebra Λ is 0-Iwanaga-Gorenstein if and only if Λ is *self-injective*, that is, $D\Lambda$ is projective as a Λ-module, or equivalently, as a Λ^{op}-module. For example, the group ring kG of a finite group G is self-injective.

(c) (Iyama and Wemyss [2014] and Curtis and Reiner [1981]) Let R be a CM local ring with canonical module ω and dimension d. An R-algebra Λ is called an *R-order* if it is CM as an R-module. Then an R-order Λ is d-Iwanaga-Gorenstein if and only if Λ is a *Gorenstein order*, i.e. $\operatorname{Hom}_R(\Lambda, \omega)$ is projective as a Λ-module, or equivalently, as a Λ^{op}-module.

An R-order Λ is called *non-singular* if $\operatorname{gl.dim} \Lambda = d$. They are classical objects for the case $d = 0, 1$ (Curtis and Reiner [1981]), and studied for $d = 2$ (Reiten and Van den Bergh [1989]). Non-singular orders are closely related to cluster tilting explained in Section 3.2.

2.3 The triangulated category of Cohen-Macaulay modules. CM modules can be defined naturally also for Iwanaga-Gorenstein rings.

Definition 2.4. Let Λ be an Iwanaga-Gorenstein ring. We call $M \in \operatorname{mod}\Lambda$ (maximal) *Cohen-Macaulay* (or *CM*) if $\operatorname{Ext}^i_\Lambda(M,\Lambda) = 0$ holds for all $i > 0$. We denote by $\operatorname{CM}\Lambda$ the category of CM Λ-modules.

We also deal with graded rings and modules. For an abelian group G and a G-graded Iwanaga-Gorenstein ring Λ, we denote by $\operatorname{CM}^G \Lambda$ the full subcategory of $\operatorname{mod}^G \Lambda$ consisting of all X which belong to $\operatorname{CM}\Lambda$ as ungraded Λ-modules.

When Λ is commutative Gorenstein, Definition 2.4 is one of the well-known equivalent conditions of CM modules. Note that, in a context of Gorenstein homological algebra (Auslander and Bridger [1969] and Enochs and Jenda [2000]), CM modules are also called Gorenstein projective, Gorenstein dimension zero, or totally reflexive.

Example 2.5. (a) Let Λ be a noetherian ring with $\operatorname{gl.dim}\Lambda < \infty$. Then $\operatorname{CM}\Lambda = \operatorname{proj}\Lambda$.

(b) Let Λ be a finite dimensional self-injective k-algebra. Then $\operatorname{CM}\Lambda = \operatorname{mod}\Lambda$.

(c) Let Λ be a Gorenstein R-order in Example 2.3(c). Then CM Λ-modules are precisely Λ-modules that are CM as R-modules.

We study the category $\operatorname{CM}^G \Lambda$ from the point of view of triangulated categories. We start with Quillen's exact categories (see Bühler [2010] for a more axiomatic definition).

Definition 2.6 (Happel [1988]). (a) An *exact category* is a full subcategory $\mathcal{F}$ of an abelian category $\mathcal{A}$ such that, for each exact sequence $0 \to X \to Y \to Z \to 0$ in $\mathcal{A}$ with $X, Z \in \mathcal{F}$, we have $Y \in \mathcal{F}$. In this case, we say that $X \in \mathcal{F}$ is *projective* if $\operatorname{Ext}^1_{\mathcal{A}}(X,\mathcal{F}) = 0$ holds. Similarly we define *injective* objects in $\mathcal{F}$.

(b) An exact category $\mathcal{F}$ in $\mathcal{A}$ is called *Frobenius* if:

- an object in $\mathcal{F}$ is projective if and only if it is injective,
- any $X \in \mathcal{F}$ admits exact sequences $0 \to Y \to P \to X \to 0$ and $0 \to X \to I \to Z \to 0$ in $\mathcal{A}$ such that P and I are projective in $\mathcal{F}$ and $Y, Z \in \mathcal{F}$.

(c) The *stable category* $\underline{\mathcal{F}}$ has the same objects as $\mathcal{F}$, and the morphisms are given by $\underline{\operatorname{Hom}}_{\mathcal{F}}(X,Y) = \operatorname{Hom}_{\mathcal{F}}(X,Y)/P(X,Y)$, where $P(X,Y)$ is the subgroup consisting of morphisms which factor through projective objects in $\mathcal{F}$.

Frobenius categories are ubiquitous in algebra. Here we give two examples.

Example 2.7. (a) For a G-graded Iwanaga-Gorenstein ring Λ, the category $\mathrm{CM}^G \Lambda$ of G-graded Cohen-Macaulay Λ-modules is a Frobenius category.

(b) For an additive category $\mathcal{A}$, the category $\mathrm{C}(\mathcal{A})$ of chain complexes in $\mathcal{A}$ is a Frobenius category, whose stable category is the homotopy category $\mathrm{K}(\mathcal{A})$.

A *triangulated category* is a triple of an additive category $\mathcal{T}$, an autoequivalence $[1]\colon \mathcal{T} \to \mathcal{T}$ (called *suspension*) and a class of diagrams $X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} X[1]$ (called *triangles*) satisfying a certain set of axioms. There are natural notions of functors and equivalences between triangulated categories, called *triangle functors* and *triangle equivalences*. For details, see e.g. Happel [1988] and Neeman [2001]. Typical examples of triangulated categories are given by the homotopy category $\mathrm{K}(\mathcal{A})$ of an additive category $\mathcal{A}$ and the derived category $\mathrm{D}(\mathcal{A})$ of an abelian category $\mathcal{A}$.

A standard construction of triangulated categories is given by the following.

Theorem 2.8 (Happel [1988]). *The stable category $\underline{\mathcal{F}}$ of a Frobenius category $\mathcal{F}$ has a canonical structure of a triangulated category.*

Such a triangulated category is called *algebraic*. Note that the suspension functor $[1]$ of $\underline{\mathcal{F}}$ is given by the cosyzygy. Thus the i-th suspension $[i]$ is the i-th cosyzygy for $i \geq 0$, and the $(-i)$-th syzygy for $i < 0$. We omit other details.

As a summary, we obtain the following.

Corollary 2.9. *Let G be an abelian group and Λ a G-graded Iwanaga-Gorenstein ring. Then $\mathrm{CM}^G \Lambda$ is a Frobenius category, and therefore the stable category $\underline{\mathrm{CM}}^G \Lambda$ has a canonical structure of a triangulated category.*

We denote by $\mathrm{D}^{\mathrm{b}}(\mathrm{mod}^G \Lambda)$ the bounded derived category of $\mathrm{mod}^G \Lambda$, and by $\mathrm{K}^{\mathrm{b}}(\mathrm{proj}^G \Lambda)$ the bounded homotopy category of $\mathrm{proj}^G \Lambda$. We regard $\mathrm{K}^{\mathrm{b}}(\mathrm{proj}^G \Lambda)$ as a thick subcategory of $\mathrm{D}^{\mathrm{b}}(\mathrm{mod}^G \Lambda)$. The *stable derived category* (Buchweitz [1987]) or the *singularity category* (Orlov [2009]) is defined as the Verdier quotient

$$\mathrm{D}^G_{\mathrm{sg}}(\Lambda) = \mathrm{D}^{\mathrm{b}}(\mathrm{mod}^G \Lambda)/\mathrm{K}^{\mathrm{b}}(\mathrm{proj}^G \Lambda).$$

This is enhanced by the Frobenius category $\mathrm{CM}^G \Lambda$ as the following result shows.

Theorem 2.10 (Buchweitz [1987], Rickard [1989a], and Keller and Vossieck [1987]). *Let G be an abelian group and Λ a G-graded Iwanaga-Gorenstein ring. Then there is a triangle equivalence $\mathrm{D}^G_{\mathrm{sg}}(\Lambda) \simeq \underline{\mathrm{CM}}^G \Lambda$.*

Let us recall the following notion (Bruns and Herzog [1993]).

Definition 2.11. Let G be an abelian group and R a G-graded Gorenstein ring with $\dim R = d$ such that $R_0 = k$ is a field and $\bigoplus_{i \neq 0} R_i$ is an ideal of R. The *a-invariant* $a \in G$ (or *Gorenstein parameter* $-a \in G$) is an element satisfying $\operatorname{Ext}^d_R(k, R(a)) \simeq k$ in $\operatorname{mod}^{\mathbb{Z}} R$.

For a G-graded noetherian ring Λ, let

$$\operatorname{qgr} \Lambda = \operatorname{mod}^G \Lambda / \operatorname{mod}^G_0 \Lambda \tag{2-1}$$

be the Serre quotient of $\operatorname{mod}^G \Lambda$ by the subcategory $\operatorname{mod}^G_0 \Lambda$ of G-graded Λ-modules of finite length (Artin and Zhang [1994]). This is classical in projective geometry. In fact, for a $\mathbb{Z}$-graded commutative noetherian ring R generated in degree 1, $\operatorname{qgr} R$ is the category $\operatorname{coh} X$ of coherent sheaves on the scheme $X = \operatorname{Proj} R$ (Serre [1955]).

The following result realizes $\mathrm{D}^{\mathbb{Z}}_{\mathrm{sg}}(R)$ and $\mathrm{D}^{\mathrm{b}}(\operatorname{qgr} R)$ inside of $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\mathbb{Z}} R)$, where $\operatorname{mod}^{\geq n} R$ is the full subcategory of $\operatorname{mod}^{\mathbb{Z}} R$ consisting of all X satisfying $X = \bigoplus_{i \geq n} X_i$, and $(-)^*$ is the duality $\mathbf{R}\operatorname{Hom}_R(-, R) \colon \mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\mathbb{Z}} R) \to \mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\mathbb{Z}} R)$.

Theorem 2.12 (Orlov [2009] and Iyama and Yang [2017])**.** *Let $R = \bigoplus_{i \geq 0} R_i$ be a $\mathbb{Z}$-graded Gorenstein ring such that R_0 is a field, and a the a-invariant of R.*

(a) *There is a triangle equivalence* $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\geq 0} R) \cap \mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\geq 1} R)^* \simeq \mathrm{D}^{\mathbb{Z}}_{\mathrm{sg}}(R)$.

(b) *There is a triangle equivalence* $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\geq 0} R) \cap \mathrm{D}^{\mathrm{b}}(\operatorname{mod}^{\geq a+1} R)^* \simeq \mathrm{D}^{\mathrm{b}}(\operatorname{qgr} R)$.

Therefore if $a = 0$, then $\mathrm{D}^{\mathbb{Z}}_{\mathrm{sg}}(R) \simeq \mathrm{D}^{\mathrm{b}}(\operatorname{qgr} R)$. If $a < 0$ (resp. $a > 0$), then there is a fully faithful triangle functor $\mathrm{D}^{\mathbb{Z}}_{\mathrm{sg}}(R) \to \mathrm{D}^{\mathrm{b}}(\operatorname{qgr} R)$ (resp. $\mathrm{D}^{\mathrm{b}}(\operatorname{qgr} R) \to \mathrm{D}^{\mathbb{Z}}_{\mathrm{sg}}(R)$). This gives a new connection between CM representations and algebraic geometry.

2.4 Representation theory.

We start with Auslander-Reiten theory.

Let R be a commutative ring, and D the Matlis duality. A triangulated category $\mathcal{T}$ is called *R-linear* if each morphism set $\operatorname{Hom}_{\mathcal{T}}(X, Y)$ has an R-module structure and the composition $\operatorname{Hom}_{\mathcal{T}}(X, Y) \times \operatorname{Hom}_{\mathcal{T}}(Y, Z) \to \operatorname{Hom}_{\mathcal{T}}(X, Z)$ is R-bilinear. It is called *Hom-finite* if each morphism set has finite length as an R-module.

Definition 2.13 (Reiten and Van den Bergh [2002])**.** A *Serre functor* is an R-linear autoequivalence $\mathbb{S} \colon \mathcal{T} \to \mathcal{T}$ such that there exists a functorial isomorphism $\operatorname{Hom}_{\mathcal{T}}(X, Y) \simeq D\operatorname{Hom}_{\mathcal{T}}(Y, \mathbb{S}X)$ for any $X, Y \in \mathcal{T}$ (called *Auslander-Reiten duality* or *Serre duality*). The composition $\tau = \mathbb{S} \circ [-1]$ is called the *AR translation*.

For $d \in \mathbb{Z}$, we say that $\mathcal{T}$ is *d-Calabi-Yau* if $[d]$ gives a Serre functor of $\mathcal{T}$.

A typical example of a Serre functor is given by a smooth projective variety X over a field k. In this case, $\mathrm{D}^{\mathrm{b}}(\operatorname{coh} X)$ has a Serre functor $- \otimes_X \omega[d]$, where ω is the canonical bundle of X and d is the dimension of X (Huybrechts [2006]).

Example 2.14 (Happel [1988] and Buchweitz, Iyama, and Yamaura [2018]). Let Λ be a finite dimensional k-algebra. Then $\mathrm{K}^{\mathrm{b}}(\operatorname{proj}\Lambda)$ has a Serre functor if and only if Λ is Iwanaga-Gorenstein, and $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda)$ has a Serre functor if and only if $\operatorname{gl.dim}\Lambda < \infty$. In both cases, the Serre functor is given by $\nu = - \overset{\mathbf{L}}{\otimes}_\Lambda (D\Lambda)$, and the AR translation is given by $\tau = \nu \circ [-1]$.

For AR theory of CM modules, we need the following notion.

Definition 2.15. Let R be a Gorenstein ring with $\dim R = d$. We denote by $\mathrm{CM}_0 R$ the full subcategory of $\mathrm{CM} R$ consisting of all X such that $X_\mathfrak{p} \in \operatorname{proj} R_\mathfrak{p}$ holds for all $\mathfrak{p} \in \operatorname{Spec} R$ with $\dim R_\mathfrak{p} < d$. When R is local, such an X is called *locally free on the punctured spectrum* (Yoshino [1990]). If R is G-graded, we denote by $\mathrm{CM}_0^G R$ the full subcategory of $\mathrm{CM}^G R$ consisting of all X which belong to $\mathrm{CM}_0 R$ as ungraded R-modules.

As before, $\mathrm{CM}_0^G R$ is a Frobenius category, and $\underline{\mathrm{CM}}_0^G R$ is a triangulated category. Note that $\mathrm{CM}_0 R = \mathrm{CM} R$ holds if and only if R satisfies Serre's (R_{d-1}) condition (i.e. $R_\mathfrak{p}$ is regular for all $\mathfrak{p} \in \operatorname{Spec} R$ with $\dim R_\mathfrak{p} < d$). This means that R has at worst an isolated singularity if R is local.

The following is a fundamental theorem of CM representations.

Theorem 2.16 (Auslander [1978] and Auslander and Reiten [1987]). *Let R be a Gorenstein ring with* $\dim R = d$. *Then $\underline{\mathrm{CM}}_0 R$ is a $(d-1)$-Calabi-Yau triangulated category. If R is G-graded and has an a-invariant $a \in G$, then $\underline{\mathrm{CM}}_0^G R$ has a Serre functor $(a)[d-1]$.*

Let us introduce a key notion in Auslander-Reiten theory. We call an additive category $\mathcal{C}$ *Krull-Schmidt* if any object in $\mathcal{C}$ is isomorphic to a finite direct sum of objects whose endomorphism rings are local. We denote by $\operatorname{ind}\mathcal{C}$ the set of isomorphism classes of indecomposable objects in $\mathcal{C}$.

Definition 2.17 (Happel [1988]). Let $\mathcal{T}$ be a Krull-Schmidt triangulated category. We call a triangle $X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} X[1]$ in $\mathcal{T}$ an *almost split triangle* if:

- X and Z are indecomposable, and $h \neq 0$ (i.e. the triangle does not split).
- Any morphism $W \to Z$ which is not a split epimorphism factors through g.
- Any morphism $X \to W$ which is not a split monomorphism factors through f.

We say that $\mathcal{T}$ *has almost split triangles* if for any indecomposable object X (resp. Z), there is an almost split triangle $X \to Y \to Z \to X[1]$.

There is a close connection between almost split triangles and Serre functors.

Theorem 2.18 (Reiten and Van den Bergh [2002]). *Let $\mathcal{T}$ be an R-linear Hom-finite Krull-Schmidt triangulated category. Then $\mathcal{T}$ has a Serre functor if and only if $\mathcal{T}$ has almost split triangles. In this case, $X \simeq \tau Z$ holds in each almost split triangle $X \to Y \to Z \to X[1]$ in $\mathcal{T}$.*

When $\mathcal{T}$ has almost split triangles, one can define the *AR quiver* of $\mathcal{T}$, which has $\operatorname{ind}\mathcal{T}$ as the set of vertices. It describes the structure of $\mathcal{T}$ (see Happel [1988]). Similarly, *almost split sequences* and the AR quiver are defined for exact categories (Assem, Simson, and Skowroński [2006] and Leuschke and Wiegand [2012]).

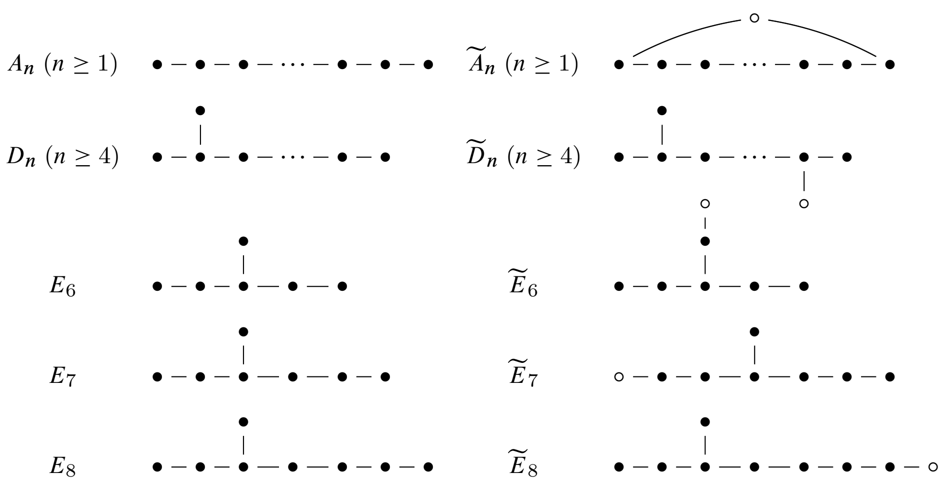
In the rest of this section, we discuss the following notion.

Definition 2.19. A finite dimensional k-algebra Λ is called *representation-finite* if $\operatorname{ind}(\operatorname{mod}\Lambda)$ is a finite set. It is also said to be *of finite representation type*.

The classification of representation-finite algebras was one of the main subjects in the 1980s. Here we recall only one theorem, and refer to Gabriel and Roĭter [1997] for further results.

A *Dynkin quiver* (resp. *extended Dynkin quiver*) is a quiver obtained by orienting each edge of one of the following diagrams A_n, D_n and E_n (resp. $\widetilde{A}_n$, $\widetilde{D}_n$ and $\widetilde{E}_n$).

(2-2)



Now we are able to state Gabriel's Theorem below. For results in non-Dynkin case, we refer to Kac's theorem in Gabriel and Roĭter [ibid.].

Theorem 2.20 (Assem, Simson, and Skowroński [2006]). *Let Q be a connected acyclic quiver and k a field. Then kQ is representation-finite if and only if Q is Dynkin. In this case, there is a bijection between $\operatorname{ind}(\operatorname{mod} kQ)$ and the set Φ_+ of positive roots in the root system of Q.*

For a quiver Q, we define a new quiver $\mathbb{Z}Q$: The set of vertices is $\mathbb{Z} \times Q_0$. The arrows are $(\ell, a)\colon (\ell, s(a)) \to (\ell, t(a))$ and $(\ell, a^*)\colon (\ell, t(a)) \to (\ell+1, s(a))$ for each $\ell \in \mathbb{Z}$ and $a \in Q_1$. For example, if $Q = [1 \xrightarrow{a} 2 \xrightarrow{b} 3]$, then $\mathbb{Z}Q$ is as follows.

$$\begin{array}{ccccccccccccccc}
 & (-2,3) & & (-1,3) & & (0,3) & & (1,3) & & (2,3) & & (3,3) & & & \\
 & & \searrow & \nearrow & \searrow & \nearrow & \searrow & \nearrow & \searrow & \nearrow & \searrow & \nearrow & & & \\
\cdots & & (-1,2) & & (0,2) & & (1,2) & & (2,2) & & (3,2) & & & \cdots & \\
 & \nearrow & \searrow & \nearrow & \searrow & \nearrow & \searrow & \nearrow & \searrow & \nearrow & \searrow & & & & \\
 & (-1,1) & & (0,1) & & (1,1) & & (2,1) & & (3,1) & & (4,1) & & &
\end{array}$$

If the underlying graph Δ of Q is a tree, then $\mathbb{Z}Q$ depends only on Δ. Thus $\mathbb{Z}Q$ is written as $\mathbb{Z}\Delta$.

The AR quiver of $\mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, kQ)$ has a simple description (Happel [1988]).

Proposition 2.21 (Happel [ibid.]). (a) *Let Λ be a finite dimensional hereditary algebra. Then there is a bijection* $\mathrm{ind}(\mathrm{mod}\,\Lambda) \times \mathbb{Z} \to \mathrm{ind}\,\mathrm{D}^{\mathrm{b}}(\mathrm{mod}\,\Lambda)$ *given by* $(X, i) \mapsto X[i]$.

(b) *For each Dynkin quiver Q, the AR quiver of* $\mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, kQ)$ *is* $\mathbb{Z}Q^{\mathrm{op}}$. *Moreover, the category* $\mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, kQ)$ *is presented by the quiver* $\mathbb{Z}Q^{\mathrm{op}}$ *with* mesh relations.

Note that $\mathbb{Z}Q$ has an automorphism τ given by $\tau(\ell, i) = (\ell-1, i)$ for $(\ell, i) \in \mathbb{Z} \times Q_0$, which corresponds to the AR translation.

Now we discuss CM-finiteness. For an additive category $\mathcal{C}$ and an object $M \in \mathcal{C}$, we denote by $\mathrm{add}\, M$ the full subcategory of $\mathcal{C}$ consisting of direct summands of finite direct sum of copies of M. We call M an *additive generator* of $\mathcal{C}$ if $\mathcal{C} = \mathrm{add}\, M$.

Definition 2.22. An Iwanaga-Gorenstein ring Λ is called *CM-finite* if $\mathrm{CM}\,\Lambda$ has an additive generator M. In this case, we call $\mathrm{End}_\Lambda(M)$ the *Auslander algebra*. When $\mathrm{CM}\,\Lambda$ is Krull-Schmidt, Λ is CM-finite if and only if $\mathrm{ind}(\mathrm{CM}\,\Lambda)$ is a finite set. It is also said to be *of finite CM type* or *representation-finite*.

Let us recall the classification of CM-finite Gorenstein rings given in the 1980s. Let k be an algebraically closed field of characteristic zero. A hypersurface $R = k[[x, y, z_2 \dots, z_d]]/(f)$ is called a *simple singularity* if

$$f = \begin{cases} x^{n+1} + y^2 + z_2^2 + \cdots + z_d^2 & A_n \\ x^{n-1} + xy^2 + z_2^2 + \cdots + z_d^2 & D_n \\ x^4 + y^3 + z_2^2 + \cdots + z_d^2 & E_6 \\ x^3 y + y^3 + z_2^2 + \cdots + z_d^2 & E_7 \\ x^5 + y^3 + z_2^2 + \cdots + z_d^2 & E_8. \end{cases} \tag{2-3}$$

We are able to state the following result. We refer to Leuschke and Wiegand [2012] for results in positive characteristic.

Theorem 2.23 (Buchweitz, Greuel, and Schreyer [1987] and Knörrer [1987]). *Let R be a complete local Gorenstein ring containing the residue field k, which is an algebraically closed field of characteristic zero. Then R is CM-finite if and only if it is a simple singularity.*

We will see that tilting theory explains why Dynkin quivers appear in both Theorems 2.20 and 2.23 (see Example 4.5 and Corollary 5.2 below).

Now we describe the AR quivers of simple singularities. Recall that each quiver Q gives a new quiver $\mathbb{Z}Q$. For an automorphism ϕ of $\mathbb{Z}Q$, an orbit quiver $\mathbb{Z}Q/\phi$ is naturally defined. For example, $\mathbb{Z}Q/\tau$ is the *double* $\overline{Q}$ of Q obtained by adding an inverse arrow $a^*: j \to i$ for each arrow $a: i \to j$.

Proposition 2.24 (Yoshino [1990] and Dieterich and Wiedemann [1986]). *Let R be a simple singularity with* $\dim R = d$. *Then the AR quiver of* $\underline{\mathrm{CM}}R$ *is $\mathbb{Z}\Delta/\phi$, where Δ and ϕ are given as follows.*

(a) *If d is even, then Δ is the Dynkin diagram of the same type as R, and $\phi = \tau$.*

(b) *If d is odd, then Δ and ϕ are given as follows.*

R	A_{2n-1}	A_{2n}	D_{2n}	D_{2n+1}	E_6	E_7	E_8
Δ	D_{n+1}	A_{2n}	D_{2n}	A_{4n-1}	E_6	E_7	E_8
ϕ	$\tau\iota$	$\tau^{1/2}$	τ^2	$\tau\iota$	$\tau\iota$	τ^2	τ^2

Here ι is the involution of $\mathbb{Z}\Delta$ induced by the non-trivial involution of Δ, and $\tau^{1/2}$ is the automorphism of $\mathbb{Z}A_{2n}$ satisfying $(\tau^{1/2})^2 = \tau$.

In dimension 2, simple singularities (over a sufficiently large field) have an alternative description as invariant subrings. This enables us to draw the AR quiver of the category CM R systematically.

Example 2.25 (Auslander [1986] and Leuschke and Wiegand [2012]). Let $k[[u,v]]$ be a formal power series ring over a field k and G a finite subgroup of $\mathrm{SL}_2(k)$ such that $\#G$ is non-zero in k. Then CM $S^G = \operatorname{add} S$ holds, and the Auslander algebra $\operatorname{End}_{S^G}(S)$ is isomorphic to the *skew group ring* $S * G$. This is a free S-module with basis G, and the multiplication is given by $(sg)(s'g') = sg(s')gg'$ for $s, s' \in S$ and $g, g' \in G$. Thus the AR quiver of CM S^G coincides with the *Gabriel quiver* of $S * G$, and hence with the *McKay quiver* of G, which is the double of an extended Dynkin quiver. This is called *algebraic McKay correspondence*.

On the other hand, the dual graph of the exceptional curves in the minimal resolution X of the singularity Spec S^G is a Dynkin graph. This is called *geometric McKay correspondence*. There is a geometric construction of CM S^G-modules using X (Artin and Verdier [1985]), which is a prototype of non-commutative crepant resolutions (Van den Bergh [2004b,a]).

3 Tilting and cluster tilting

3.1 Tilting theory. Tilting theory is a Morita theory for triangulated categories. It has an origin in Bernstein-Gelfand-Ponomarev reflection for quiver representations, and established by works of Brenner-Butler, Happel-Ringel, Rickard, Keller and others (see e.g. Angeleri Hügel, Happel, and Krause [2007]). The class of silting objects was introduced to complete the class of tilting objects in the study of t-structures (Keller and Vossieck [1988]) and mutation (Aihara and Iyama [2012]).

Definition 3.1. Let $\mathcal{T}$ be a triangulated category. A full subcategory of $\mathcal{T}$ is *thick* if it is closed under cones, $[\pm 1]$ and direct summands. We call an object $T \in \mathcal{T}$ *tilting* (resp. *silting*) if $\mathrm{Hom}_{\mathcal{T}}(T, T[i]) = 0$ holds for all integers $i \neq 0$ (resp. $i > 0$), and the smallest thick subcategory of $\mathcal{T}$ containing T is $\mathcal{T}$.

The principal example of tilting objects appears in $\mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\Lambda)$ for a ring Λ. It has a tilting object given by the stalk complex Λ concentrated in degree zero. Conversely, any triangulated category with a tilting object is triangle equivalent to $\mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\Lambda)$ under mild assumptions (see Kimura [2016] for a detailed proof).

Theorem 3.2 (Keller [1994]). *Let $\mathcal{T}$ be an algebraic triangulated category and $T \in \mathcal{T}$ a tilting object. If $\mathcal{T}$ is idempotent complete, then there is a triangle equivalence $\mathcal{T} \simeq \mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\mathrm{End}_{\mathcal{T}}(T))$ sending T to $\mathrm{End}_{\mathcal{T}}(T)$.*

As an application, one can deduce Rickard's fundamental Theorem (Rickard [1989b]), characterizing when two rings are derived equivalent in terms of tilting objects. Another application is the following converse of Proposition 2.21(b).

Example 3.3. Let $\mathcal{T}$ be a k-linear Hom-finite Krull-Schmidt algebraic triangulated category over an algebraically closed field k. If the AR quiver of $\mathcal{T}$ is $\mathbb{Z}Q$ for a Dynkin quiver Q, then $\mathcal{T}$ has a tilting object $T = \bigoplus_{i \in Q_0}(0, i)$ for $(0, i) \in \mathbb{Z} \times Q_0 = (\mathbb{Z}Q)_0 = \mathrm{ind}\,\mathcal{T}$. Thus there is a triangle equivalence $\mathcal{T} \simeq \mathrm{D}^{\mathrm{b}}(\mathrm{mod}\,kQ^{\mathrm{op}})$.

The following is the first main problem we will discuss in this paper.

Problem 3.4. *Find a G-graded Iwanaga-Gorenstein ring Λ such that there is a triangle equivalence $\underline{\mathrm{CM}}^G\Lambda \simeq \mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\Gamma)$ for some ring Γ. Equivalently (by Theorem 3.2), find a G-graded Iwanaga-Gorenstein ring Λ such that there is a tilting object in $\underline{\mathrm{CM}}^G\Lambda$.*

3.2 Cluster tilting and higher Auslander-Reiten theory. The notion of cluster tilting appeared naturally in a context of higher Auslander-Reiten theory (Iyama [2008]). It also played a central role in categorification of cluster algebras (Fomin and Zelevinsky [2002]) by using cluster categories, a new class of triangulated categories introduced in Buan, Marsh, Reineke, Reiten, and Todorov [2006], and preprojective algebras (Geiss, Leclerc, and Schröer [2013]). Here we explain only the minimum necessary background for the aim of this paper.

Let Λ be a finite dimensional k-algebra with $\operatorname{gl.dim}\Lambda \leq d$. Then $\mathrm{D^b}(\operatorname{mod}\Lambda)$ has a Serre functor ν by Example 2.14. Using the *higher AR translation* $\nu_d := \nu \circ [-d]$ of $\mathrm{D^b}(\operatorname{mod}\Lambda)$, the *orbit category* $\mathrm{C}^{\circ}_d(\Lambda) = \mathrm{D^b}(\operatorname{mod}\Lambda)/\nu_d$ is defined. It has the same objects as $\mathrm{D^b}(\operatorname{mod}\Lambda)$, and the morphism space is given by

$$\operatorname{Hom}_{\mathrm{C}^{\circ}_d(\Lambda)}(X,Y) = \bigoplus_{i\in\mathbb{Z}} \operatorname{Hom}_{\mathrm{D^b}(\operatorname{mod}\Lambda)}(X, \nu_d^i(Y)),$$

where the composition is defined naturally. In general, $\mathrm{C}^{\circ}_d(\Lambda)$ does not have a natural structure of a triangulated category. The *d-cluster category* of Λ is a triangulated category $\mathrm{C}_d(\Lambda)$ containing $\mathrm{C}^{\circ}_d(\Lambda)$ as a full subcategory such that the composition $\mathrm{D^b}(\operatorname{mod}\Lambda) \to \mathrm{C}^{\circ}_d(\Lambda) \subset \mathrm{C}_d(\Lambda)$ is a triangle functor. It was constructed in Buan, Marsh, Reineke, Reiten, and Todorov [2006] for hereditary case where $\mathrm{C}_d(\Lambda) = \mathrm{C}^{\circ}_d(\Lambda)$ holds, and in Keller [2005, 2011], Amiot [2009], and Guo [2011] for general case by using a DG enhancement of $\mathrm{D^b}(\operatorname{mod}\Lambda)$.

We say that Λ is ν_d*-finite* if $H^0(\nu_d^{-i}(\Lambda)) = 0$ holds for $i \gg 0$. This is automatic if $\operatorname{gl.dim}\Lambda < d$. In the hereditary case $d = 1$, Λ is ν_1-finite if and only if it is representation-finite. The following is a basic property of d-cluster categories.

Theorem 3.5 (Amiot [2009] and Guo [2011])**.** *Let Λ be a finite dimensional k-algebra with* $\operatorname{gl.dim}\Lambda \leq d$. *Then Λ is ν_d-finite if and only if $\mathrm{C}_d(\Lambda)$ is Hom-finite. In this case, $\mathrm{C}_d(\Lambda)$ is a d-Calabi-Yau triangulated category.*

Thus, if Λ is ν_d-finite, then $\mathrm{C}_d(\Lambda)$ never has a tilting object. But the object Λ in $\mathrm{C}_d(\Lambda)$ still enjoys a similar property to tilting objects. Now we recall the following notion, introduced in Iyama [2007b] as a *maximal* $(d-1)$*-orthogonal* subcategory.

Definition 3.6 (Iyama [ibid.])**.** Let $\mathcal{T}$ be a triangulated or exact category and $d \geq 1$. We call a full subcategory $\mathcal{C}$ of $\mathcal{T}$ *d-cluster tilting* if $\mathcal{C}$ is a functorially finite subcategory of $\mathcal{T}$ such that

$$\begin{aligned}\mathcal{C} &= \{X \in \mathcal{T} \mid \forall i \in \{1,2,\dots,d-1\}\ \operatorname{Ext}^i_{\mathcal{T}}(\mathcal{C}, X) = 0\}\\ &= \{X \in \mathcal{T} \mid \forall i \in \{1,2,\dots,d-1\}\ \operatorname{Ext}^i_{\mathcal{T}}(X, \mathcal{C}) = 0\}.\end{aligned}$$

We call an object $T \in \mathcal{T}$ *d-cluster tilting* if $\operatorname{add} T$ is a d-cluster tilting subcategory.

If $\mathcal{T}$ has a Serre functor $\mathbb{S}$, then it is easy to show $(\mathbb{S}\circ[-d])(\mathcal{C}) = \mathcal{C}$. Thus it is natural in our setting $\mathcal{T} = \mathrm{D^b}(\operatorname{mod}\Lambda)$ to consider the full subcategory

$$\mathrm{U}_d(\Lambda) := \operatorname{add}\{\nu_d^i(\Lambda) \mid i \in \mathbb{Z}\} \subset \mathrm{D^b}(\operatorname{mod}\Lambda). \tag{3-1}$$

Equivalently, $\mathrm{U}_d(\Lambda) = \pi^{-1}(\operatorname{add} \pi\Lambda)$ for the functor $\pi : \mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda) \to \mathrm{C}_d(\Lambda)$. In the hereditary case $d = 1$, $\mathrm{U}_1(\Lambda) = \mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda)$ holds if Λ is representation-finite, and otherwise $\mathrm{U}_1(\Lambda)$ is the connected component of the AR quiver of $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda)$ containing Λ. This observation is generalized as follows.

Theorem 3.7 (Amiot [2009] and Iyama [2011]). *Let Λ be a finite dimensional k-algebra with* $\operatorname{gl.dim}\Lambda \leq d$. *If Λ is ν_d-finite, then $\mathrm{C}_d(\Lambda)$ has a d-cluster tilting object Λ, and* $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda)$ *has a d-cluster tilting subcategory* $\mathrm{U}_d(\Lambda)$.

We define a full subcategory of $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda)$ by

$$\mathrm{D}^{d\mathbb{Z}}(\operatorname{mod}\Lambda) = \{X \in \mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda) \mid \forall i \in \mathbb{Z} \setminus d\mathbb{Z},\ H^i(X) = 0\}.$$

If $\operatorname{gl.dim}\Lambda \leq d$, then any object in $\mathrm{D}^{d\mathbb{Z}}(\operatorname{mod}\Lambda)$ is isomorphic to a finite direct sum of $X[di]$ for some $X \in \operatorname{mod}\Lambda$ and $i \in \mathbb{Z}$. This generalizes Proposition 2.21(a) for hereditary algebras, and motivates the following definition.

Definition 3.8 (Herschend, Iyama, and Oppermann [2014]). Let $d \geq 1$. A finite dimensional k-algebra Λ is called *d-hereditary* if $\operatorname{gl.dim}\Lambda \leq d$ and $\mathrm{U}_d(\Lambda) \subset \mathrm{D}^{d\mathbb{Z}}(\operatorname{mod}\Lambda)$.

It is clear that 1-hereditary algebras are precisely hereditary algebras. We have the following dichotomy of d-hereditary algebras.

Theorem 3.9 (Herschend, Iyama, and Oppermann [ibid.]). *Let Λ be a ring-indecomposable finite dimensional k-algebra with* $\operatorname{gl.dim}\Lambda \leq d$. *Then Λ is d-hereditary if and only if either (i) or (ii) holds:*

(i) *There exists a d-cluster tilting object in* $\operatorname{mod}\Lambda$.

(ii) $\nu_d^{-i}(\Lambda) \in \operatorname{mod}\Lambda$ *holds for any* $i \geq 0$.

When $d = 1$, the above (i) holds if and only if Λ is representation-finite, and the above (ii) holds if and only if Λ is d-representation-infinite.

Definition 3.10. Let Λ be a d-hereditary algebra. We call Λ *d-representation-finite* if the above (i) holds, and *d-representation-infinite* if the above (ii) holds.

Example 3.11. (a) Let $\Lambda = kQ$ for a connected acyclic quiver Q. Then Λ is 1-representation-finite if Q is Dynkin, and 1-representation-infinite otherwise.

(b) Let X be a smooth projective variety with $\dim X = d$, and $T \in \operatorname{coh} X$ a tilting object in $\mathrm{D}^{\mathrm{b}}(\operatorname{coh} X)$. Then $\Lambda = \operatorname{End}_X(T)$ always satisfies $\operatorname{gl.dim}\Lambda \geq d$. If the equality holds, then Λ is d-representation-infinite (Buchweitz and Hille [2014]).

(c) There is a class of finite dimensional k-algebras called *Fano algebras* (Minamoto [2012] and Minamoto and Mori [2011]) in non-commutative algebraic geometry. So-called *extremely Fano algebras* Λ with $\operatorname{gl.dim} \Lambda = d$ are d-representation-infinite.

It is known that d-cluster tilting subcategories of a triangulated (resp. exact) category $\mathcal{T}$ enjoy various properties which should be regarded as higher analogs of those of $\mathcal{T}$. For example, they have *almost split* $(d+2)$*-angles* by Iyama and Yoshino [2008] (resp. d*-almost split sequences* by Iyama [2007b]), and structures of $(d+2)$*-angulated categories* by Geiss, Keller, and Oppermann [2013] (resp. d*-abelian categories* by Jasso [2016]). These motivate the following definition.

Definition 3.12 (cf. Definition 2.22)**.** An Iwanaga-Gorenstein ring Λ is called d*-CM-finite* if there exists a d-cluster tilting object M in $\operatorname{CM} \Lambda$. In this case, we call $\operatorname{End}_\Lambda(M)$ the d*-Auslander algebra* and $\underline{\operatorname{End}}_\Lambda(M)$ the *stable* d*-Auslander algebra*.

1-CM-finiteness coincides with classical CM-finiteness since 1-cluster tilting objects are precisely additive generators. d*-Auslander correspondence* gives a characterization of a certain nice class of algebras with finite global dimension as d-Auslander algebras (Iyama [2007a]). As a special case, it gives a connection with non-commutative crepant resolutions (NCCRs) of Van den Bergh [2004a]. Recall that a reflexive module M over a Gorenstein ring R gives an *NCCR* $\operatorname{End}_R(M)$ of R if $\operatorname{End}_R(M)$ is a non-singular R-order (see Example 2.3(c)).

Theorem 3.13 (Iyama [2007a])**.** *Let R be a Gorenstein ring with* $\dim R = d+1$. *Assume $M \in$* CM R *has R as a direct summand. Then M is a d-cluster tilting object in* CM R *if and only if M gives an NCCR of R and R satisfies Serre's (R_d) condition.*

The following generalizes Example 2.25.

Example 3.14 (Iyama [2007b] and Van den Bergh [2004a])**.** Let $S = k[[x_0, \dots, x_d]]$ be a formal power series ring and G a finite subgroup of $\mathrm{SL}_{d+1}(k)$ such that $\#G$ is non-zero in k. Then the S^G-module S gives an NCCR $\operatorname{End}_{S^G}(S) = S * G$ of S^G. If S^G has at worst an isolated singularity, then S is a d-cluster tilting object in $\operatorname{CM} S^G$, and hence S^G is d-CM-finite with the d-Auslander algebra $S * G$. As in Example 2.25, the quiver of $\operatorname{add} S$ coincides with the Gabriel quiver of $S * G$ and with the McKay quiver of G.

The following is the second main problem we will discuss in this paper.

Problem 3.15. *Find a d-CM-finite Iwanaga-Gorenstein ring. More strongly (by Theorem 3.7), find an Iwanaga-Gorenstein ring Λ such that there is a triangle equivalence* $\underline{\operatorname{CM}}\Lambda \simeq \mathrm{C}_d(\Gamma)$ *for some algebra* Γ.

We refer to Erdmann and Holm [2008] and Bergh [2014] for some necessary conditions for d-CM-finiteness. Besides results in this paper, a number of examples of NCCRs have been found, see e.g. Leuschke [2012], Wemyss [2016], and Špenko and Van den Bergh [2017] and references therein.

It is natural to ask how the notion of d-CM-finiteness is related to CM-tameness (e.g. Burban and Y. Drozd [2008]) and also the representation type of homogeneous coordinate rings of projective varieties (e.g. Faenzi and Malaspina [2017]).

4 Results in dimension 0 and 1

4.1 Dimension zero. In this subsection, we consider finite dimensional Iwanaga-Gorenstein algebras. We start with a classical result due to Happel [1988]. Let Λ be a finite dimensional k-algebra. The *trivial extension algebra* of Λ is $T(\Lambda) = \Lambda \oplus D\Lambda$, where the multiplication is given by $(\lambda, f)(\lambda', f') = (\lambda\lambda', \lambda f' + f\lambda')$ for $(\lambda, f), (\lambda', f') \in T(\Lambda)$. This is clearly a self-injective k-algebra, and has a $\mathbb{Z}$-grading given by $T(\Lambda)_0 = \Lambda$, $T(\Lambda)_1 = D\Lambda$ and $T(\Lambda)_i = 0$ for $i \neq 0, 1$.

Theorem 4.1 (Happel [ibid.]). *Let Λ be a finite dimensional k-algebra with* $\mathrm{gl.dim}\,\Lambda < \infty$. *Then $\underline{\mathrm{mod}}^{\mathbb{Z}}\,T(\Lambda)$ has a tilting object Λ such that $\underline{\mathrm{End}}^{\mathbb{Z}}_{T(\Lambda)}(\Lambda) \simeq \Lambda$, and there is a triangle equivalence*

$$\underline{\mathrm{mod}}^{\mathbb{Z}}\,T(\Lambda) \simeq \mathrm{D^b}(\mathrm{mod}\,\Lambda). \tag{4-1}$$

As an application, it follows from Gabriel's Theorem 2.20 and covering theory that $T(kQ)$ is representation-finite for any Dynkin quiver Q. More generally, a large family of representation-finite self-injective algebras was constructed from Theorem 4.1. See a survey article (Skowroński [2006]).

Recently, Theorem 4.1 was generalized to a large class of $\mathbb{Z}$-graded self-injective algebras Λ. For $X \in \mathrm{mod}^{\mathbb{Z}}\,\Lambda$, let $X_{\geq 0} = \bigoplus_{i \geq 0} X_i$.

Theorem 4.2 (Yamaura [2013]). *Let $\Lambda = \bigoplus_{i\geq 0} \Lambda_i$ be a $\mathbb{Z}$-graded finite dimensional self-injective k-algebra such that* $\mathrm{gl.dim}\,\Lambda_0 < \infty$. *Then $\underline{\mathrm{mod}}^{\mathbb{Z}}\,\Lambda$ has a tilting object $T = \bigoplus_{i>0} \Lambda(i)_{\geq 0}$, and there is a triangle equivalence $\underline{\mathrm{mod}}^{\mathbb{Z}}\,\Lambda \simeq \mathrm{K^b}(\mathrm{proj}\,\underline{\mathrm{End}}^{\mathbb{Z}}_{\Lambda}(T))$.*

If $\mathrm{soc}\,\Lambda \subset \Lambda_a$ for some $a \in \mathbb{Z}$, then $\underline{\mathrm{End}}^{\mathbb{Z}}_{\Lambda}(T)$ has a simple description

$$\underline{\mathrm{End}}^{\mathbb{Z}}_{\Lambda}(T) \simeq \begin{bmatrix} \Lambda_0 & 0 & \cdots & 0 & 0 \\ \Lambda_1 & \Lambda_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Lambda_{a-2} & \Lambda_{a-3} & \cdots & \Lambda_0 & 0 \\ \Lambda_{a-1} & \Lambda_{a-2} & \cdots & \Lambda_1 & \Lambda_0 \end{bmatrix}.$$

For example, if $\Lambda = k[x]/(x^{a+1})$ with $\deg x = 1$, then $\underline{\mathrm{End}}^{\mathbb{Z}}_{\Lambda}(T)$ is the path algebra $k\mathbb{A}_a$ of the quiver of type A_a.

We end this subsection with posing the following open problem.

Problem 4.3. *Let $\Lambda = \bigoplus_{i\geq 0} \Lambda_i$ be a $\mathbb{Z}$-graded finite dimensional Iwanaga-Gorenstein algebra. When does $\underline{\mathrm{CM}}^{\mathbb{Z}}\Lambda$ have a tilting object?*

Recently, it was shown in Lu and Zhu [2017] and Kimura, Minamoto, and Yamaura [n.d.] independently that if $\Lambda = \bigoplus_{i\geq 0} \Lambda_i$ is a $\mathbb{Z}$-graded finite dimensional 1-Iwanaga-Gorenstein algebra satisfying $\mathrm{gl.dim}\,\Lambda_0 < \infty$, then the stable category $\underline{\mathrm{CM}}^{\mathbb{Z}}\Lambda$ has a silting object. We will see some other results in Section 5.4. We refer to Darpö and Iyama [2017] for some results on Problem 3.15.

4.2 Dimension one.

In this subsection, we consider a $\mathbb{Z}$-graded Gorenstein ring $R = \bigoplus_{i\geq 0} R_i$ with $\dim R = 1$ such that R_0 is a field. Let S be the multiplicative set of all homogeneous non-zerodivisors of R, and $K = RS^{-1}$ the $\mathbb{Z}$-graded total quotient ring. Then there exists a positive integer p such that $K(p) \simeq K$ as $\mathbb{Z}$-graded R-modules. In this setting, we have the following result (see Definitions 2.15 and 2.11 for $\mathrm{CM}_0^{\mathbb{Z}} R$ and the a-invariant).

Theorem 4.4 (Buchweitz, Iyama, and Yamaura [2018])**.** *Let $R = \bigoplus_{i\geq 0} R_i$ be a $\mathbb{Z}$-graded Gorenstein ring with $\dim R = 1$ such that R_0 is a field, and a the a-invariant of R.*

(a) *Assume $a \geq 0$. Then $\underline{\mathrm{CM}}_0^{\mathbb{Z}} R$ has a tilting object $T = \bigoplus_{i=1}^{a+p} R(i)_{\geq 0}$, and there is a triangle equivalence $\underline{\mathrm{CM}}_0^{\mathbb{Z}} R \simeq \mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\underline{\mathrm{End}}_R^{\mathbb{Z}}(T))$.*

(b) *Assume $a < 0$. Then $\underline{\mathrm{CM}}_0^{\mathbb{Z}} R$ has a silting object $\bigoplus_{i=1}^{a+p} R(i)_{\geq 0}$. Moreover, it has a tilting object if and only if R is regular.*

An important tool in the proof is Theorem 2.12. The endomorphism algebra of T above has the following description.

$$\underline{\mathrm{End}}_R^{\mathbb{Z}}(T) = \begin{bmatrix} R_0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ R_1 & R_0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ R_{a-2} & R_{a-3} & \cdots & R_0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ R_{a-1} & R_{a-2} & \cdots & R_1 & R_0 & 0 & 0 & \cdots & 0 & 0 \\ K_a & K_{a-1} & \cdots & K_2 & K_1 & K_0 & K_{-1} & \cdots & K_{2-p} & K_{1-p} \\ K_{a+1} & K_a & \cdots & K_3 & K_2 & K_1 & K_0 & \cdots & K_{3-p} & K_{2-p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ K_{a+p-2} & K_{a+p-3} & \cdots & K_p & K_{p-1} & K_{p-2} & K_{p-3} & \cdots & K_0 & K_{-1} \\ K_{a+p-1} & K_{a+p-2} & \cdots & K_{p+1} & K_p & K_{p-1} & K_{p-2} & \cdots & K_1 & K_0 \end{bmatrix}.$$

As an application, we obtain the following graded version of Proposition 2.24(b).

Example 4.5. Let $R = k[x, y]/(f)$ be a simple singularity (2-3) with $\dim R = 1$ and the grading given by the list below. Then there is a triangle equivalence $\underline{\mathrm{CM}}^{\mathbb{Z}} R \simeq \mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, kQ)$, where Q is the Dynkin quiver in the list below. In particular, the AR quiver of $\underline{\mathrm{CM}}^{\mathbb{Z}} R$ is $\mathbb{Z}Q^{\mathrm{op}}$ (Araya [1999]).

R	A_{2n-1}	A_{2n}	D_{2n}	D_{2n+1}	E_6	E_7	E_8
$(\deg x, \deg y)$	$(1, n)$	$(2, 2n+1)$	$(1, n-1)$	$(2, 2n-1)$	$(3, 4)$	$(2, 3)$	$(3, 5)$
Q	D_{n+1}	A_{2n}	D_{2n}	A_{4n-1}	E_6	E_7	E_8

This gives a conceptual proof of the classical result that simple singularities in dimension 1 are CM-finite (Jacobinski [1967], J. A. Drozd and Roĭter [1967], and Greuel and Knörrer [1985]).

In the following special case, one can construct a different tilting object, whose endomorphism algebra is 2-representation-finite (Definition 3.10). This is closely related to the 2-cluster tilting object constructed in Burban, Iyama, Keller, and Reiten [2008].

Theorem 4.6 (Herschend and Iyama [n.d.]). *Let $R = k[x, y]/(f)$ be a hypersurface singularity with $f = f_1 f_2 \cdots f_n$ for linear forms f_i and* $\deg x = \deg y = 1$. *Assume that R is reduced.*

(a) *$\underline{\mathrm{CM}}^{\mathbb{Z}} R$ has a tilting object*

$$U = \bigoplus_{i=1}^{n} (k[x, y]/(f_1 f_2 \cdots f_i) \oplus k[x, y]/(f_1 f_2 \cdots f_i)(1))$$

(b) *$\underline{\mathrm{End}}_R^{\mathbb{Z}}(U)$ is a 2-representation-finite algebra. It is the Jacobian algebra of a certain quiver with potential.*

We refer to Demonet and Luo [2016], Jensen, King, and Su [2016], and Gelinas [2017] for other results in dimension one.

5 Preprojective algebras

5.1 Classical preprojective algebras. Preprojective algebras are widely studied objects with various applications, e.g. cluster algebras (Geiss, Leclerc, and Schröer [2013]), quantum groups (Kashiwara and Y. Saito [1997] and Lusztig [1991]), quiver varieties (Nakajima [1994]). Here we discuss a connection to CM representations.

Let Q be an acyclic quiver, and $\overline{Q}$ the double of Q obtained by adding an inverse arrow $a^*: j \to i$ for each arrow $a: i \to j$ in Q. The *preprojective algebra* of Q is the factor algebra of the path algebra $k\overline{Q}$ defined by

$$\Pi = k\overline{Q}/(\sum_{a \in Q_1} (aa^* - a^*a)). \tag{5-1}$$

We regard Π as a $\mathbb{Z}$-graded algebra by $\deg a = 0$ and $\deg a^* = 1$ for any $a \in Q_1$. Clearly $\Pi_0 = kQ$ holds. Moreover $\Pi_1 = \mathrm{Ext}^1_{kQ}(D(kQ), kQ)$ as a kQ-bimodule, and Π is isomorphic to the tensor algebra $T_{kQ}\,\mathrm{Ext}^1_{kQ}(D(kQ), kQ)$. Thus the kQ-module Π_i is isomorphic to the *preprojective kQ-module* $H^0(\tau^{-i}(kQ))$, where $\tau = \nu \circ [-1]$ is the AR translation. This is the reason why Π is called the preprojective algebra. Moreover, for the category $\mathrm{U}_1(kQ)$ defined in (3-1), there is an equivalence

$$\mathrm{U}_1(kQ) = \mathrm{add}\{\tau^{-i}(kQ) \mid i \in \mathbb{Z}\} \simeq \mathrm{proj}^{\mathbb{Z}}\,\Pi \tag{5-2}$$

given by $X \mapsto \bigoplus_{i \in \mathbb{Z}} \mathrm{Hom}_{\mathrm{U}_1(kQ)}(kQ, \tau^{-i}(X))$, which gives the following trichotomy.

Q	Dynkin	extended Dynkin	else
kQ	representation-finite	representation-tame	representation-wild
$\dim_k \Pi_i$	$\dim_k \Pi < \infty$	linear growth	exponential growth

It was known in 1980s that, Π in the extended Dynkin case has a close connection to simple singularities.

Theorem 5.1 (Auslander [1978], Geigle and Lenzing [1987, 1991], and Reiten and Van den Bergh [1989]). *Let Π be a preprojective algebra of an extended Dynkin quiver Q, e the vertex $\circ$ in* (2-2)*, and $R = e\Pi e$.*

(a) *R is a simple singularity $k[x, y, z]/(f)$ in dimension 2 with induced $\mathbb{Z}$-grading below, where p in type A_n is the number of clockwise arrows in Q. (Note that f coincides with* (2-3) *after a change of variables if k is sufficiently large.)*

Q, R	f	$(\deg x, \deg y, \deg z)$
A_n	$x^{n+1} - yz$	$(1, p, n+1-p)$
D_n	$x(y^2 + x^{\ell-1}y) + z^2$ if $n = 2\ell$ $x(y^2 + x^{\ell-1}z) + z^2$ if $n = 2\ell+1$	$(2, n-2, n-1)$
E_6	$x^2z + y^3 + z^2$	$(3, 4, 6)$
E_7	$x^3y + y^3 + z^2$	$(4, 6, 9)$
E_8	$x^5 + y^3 + z^2$	$(6, 10, 15)$

(b) *Πe is an additive generator of* CM R *and satisfies* $\mathrm{End}_R(\Pi e) = \Pi$. *Therefore R is CM-finite with an Auslander algebra Π.*

(c) Π *is Morita equivalent to the skew group ring* $k[u,v]*G$ *for a finite subgroup* G *of* $\mathrm{SL}_2(k)$ *if* k *is sufficiently large (cf. Example 2.25).*

By (b) and (5-2) above, there are equivalences $\mathrm{CM}^{\mathbb{Z}} R \simeq \mathrm{proj}^{\mathbb{Z}} \Pi \simeq \mathrm{U}_1(kQ) \subset \mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, kQ)$. Thus the AR quivers of $\mathrm{CM}^{\mathbb{Z}} R$ and $\underline{\mathrm{CM}}^{\mathbb{Z}} R$ are given by $\mathbb{Z}Q^{\mathrm{op}}$ and $\mathbb{Z}(Q^{\mathrm{op}}\backslash\{e\})$ respectively. Now the following result follows from Example 3.3.

Corollary 5.2. *Under the setting in Theorem 5.1, there is a triangle equivalence* $\underline{\mathrm{CM}}^{\mathbb{Z}} R \simeq \mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, kQ/(e))$.

Two other proofs were given in Kajiura, K. Saito, and A. Takahashi [2007], one uses explicit calculations of $\mathbb{Z}$-graded matrix factorizations, and the other uses Theorem 2.12. In Theorem 5.8 below, we deduce Corollary 5.2 from a general result on higher preprojective algebras. We refer to Kajiura, K. Saito, and A. Takahashi [2009] and Lenzing and de la Peña [2011] for results for some other hypersurfaces in dimension 2.

5.2 Higher preprojective algebras. There is a natural analog of preprojective algebras for finite dimensional algebras with finite global dimension.

Definition 5.3 (Iyama and Oppermann [2013])**.** Let Λ be a finite dimensional k-algebra with $\mathrm{gl.dim}\,\Lambda \le d$. We regard the highest extension $\mathrm{Ext}^d_\Lambda(D\Lambda,\Lambda)$ as a Λ-bimodule naturally, and define the $(d+1)$-*preprojective algebra* as the tensor algebra

$$\Pi_{d+1}(\Lambda) = T_\Lambda \mathrm{Ext}^d_\Lambda(D\Lambda,\Lambda).$$

This is the 0-th cohomology of the *Calabi-Yau completion* of Λ (Keller [2011]). For example, for an acyclic quiver Q, $\Pi_2(kQ)$ is the preprojective algebra (5-1).

The algebra $\Pi = \Pi_{d+1}(\Lambda)$ has an alternative description in terms of the higher AR translation $\nu_d = \nu \circ [-d]$ of $\mathrm{D}^{\mathrm{b}}(\mathrm{mod}\,\Lambda)$. The $\mathbb{Z}$-grading on Π is given by

$$\Pi_i = \mathrm{Ext}^d_\Lambda(D\Lambda,\Lambda)^{\otimes_\Lambda i} = \mathrm{Hom}_{\mathrm{D}^{\mathrm{b}}(\mathrm{mod}\,\Lambda)}(\Lambda, \nu_d^{-i}(\Lambda))$$

for $i \ge 0$. Thus there is an isomorphism $\Pi \simeq \mathrm{End}_{\mathrm{C}_d(\Lambda)}(\Lambda)$ and an equivalence

$$\mathrm{U}_d(\Lambda) \simeq \mathrm{proj}^{\mathbb{Z}} \Pi \tag{5-3}$$

given by $X \mapsto \bigoplus_{i\in\mathbb{Z}} \mathrm{Hom}_{\mathrm{U}_d(\Lambda)}(\Lambda, \nu_d^{-i}(X))$. In particular, Π is finite dimensional if and only if Λ is ν_d-finite.

We see below that $\Pi_{d+1}(\Lambda)$ enjoys nice homological properties if Λ is d-hereditary.

Definition 5.4 (cf. Ginzburg [2006])**.** Let $\Gamma = \bigoplus_{i\ge 0}\Gamma_i$ be a $\mathbb{Z}$-graded k-algebra. We denote by $\Gamma^{\mathrm{e}} = \Gamma^{\mathrm{op}} \otimes_k \Gamma$ the enveloping algebra of Γ. We say that Γ is a d-*Calabi-Yau algebra of* a-*invariant* a (or *Gorenstein parameter* $-a$) if Γ belongs to $\mathrm{K}^{\mathrm{b}}(\mathrm{proj}^{\mathbb{Z}}\Gamma^{\mathrm{e}})$ and $\mathbf{R}\mathrm{Hom}_{\Gamma^{\mathrm{e}}}(\Gamma,\Gamma^{\mathrm{e}})(a)[d] \simeq \Gamma$ holds in $\mathrm{D}(\mathrm{Mod}^{\mathbb{Z}}\Gamma^{\mathrm{e}})$.

For example, the $\mathbb{Z}$-graded polynomial algebra $k[x_1, \dots, x_d]$ with $\deg x_i = a_i$ is a d-Calabi-Yau algebra of a-invariant $-\sum_{i=1}^{n} a_i$.

Now we give a homological characterization of the $(d+1)$-preprojective algebras of d-representation-infinite algebras (Definition 3.10) as the explicit correspondence.

Theorem 5.5 (Keller [2011], Minamoto and Mori [2011], and Amiot, Iyama, and Reiten [2015]). *There exists a bijection between the set of isomorphism classes of d-representation-infinite algebras Λ and the set of isomorphism classes of $(d+1)$-Calabi-Yau algebras Γ of a-invariant -1. It is given by $\Lambda \mapsto \Pi_{d+1}(\Lambda)$ and $\Gamma \mapsto \Gamma_0$.*

Note that Γ above is usually non-noetherian. If Γ is right graded coherent, then for the category $\operatorname{qgr}\Gamma$ defined in (2-1), there is a triangle equivalence (Minamoto [2012])

$$\mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Lambda) \simeq \mathrm{D}^{\mathrm{b}}(\operatorname{qgr}\Gamma). \tag{5-4}$$

Applying Theorem 5.5 for $d = 1, 2$, we obtain the following observations (see Van den Bergh [2015] for a structure theorem of (ungraded) Calabi-Yau algebras).

Example 5.6. Let k be an algebraically closed field.

(a) (cf. Bocklandt [2008]) 2-Calabi-Yau algebras of a-invariant -1 are precisely the preprojective algebras of disjoint unions of non-Dynkin quivers.

(b) (cf. Bocklandt [2008] and Herschend and Iyama [2011]) 3-Calabi-Yau algebras of a-invariant -1 are precisely the Jacobian algebras of quivers with 'good' potential with cuts.

The setting of our main result is the following.

Assumption 5.7. Let Γ be a $(d+1)$-Calabi-Yau algebras of a-invariant -1. Equivalently by Theorem 5.5, Γ is a $(d+1)$-preprojective algebra of some d-representation-infinite algebra. We assume that the following conditions hold for $\Lambda = \Gamma_0$.

(i) Γ is a noetherian ring, $e \in \Lambda$ is an idempotent and $\dim_k(\Gamma/(e)) < \infty$.

(ii) $e\Lambda(1-e) = 0$.

For example, let Q be an extended Dynkin quiver. If the vertex $\circ$ in (2-2) is a sink, then $\Gamma = \Pi_2(kQ)$ and $e = \circ$ satisfy Assumption 5.7 by Theorem 5.1.

Under Assumption 5.7(i), let $R = e\Gamma e$. Then R is a $(d+1)$-Iwanaga-Gorenstein ring, and the (Γ, R)-bimodule Γe plays an important role. It is a CM R-module, and gives a d-cluster tilting object in $\mathrm{CM}\, R$. Moreover the natural morphism $\Gamma \to \operatorname{End}_R(\Gamma e)$ is an isomorphism. Thus R is d-CM-finite and has a d-Auslander algebra Γ. The proof of these statements is parallel to Example 3.14.

Regarding Γe as a $\mathbb{Z}$-graded R-module, we consider the composition

$$F\colon \mathrm{D^b}(\operatorname{mod}\Lambda/(e)) \to \mathrm{D^b}(\operatorname{mod}\Lambda) \xrightarrow{-\overset{\mathbf{L}}{\otimes}_\Lambda \Gamma e} \mathrm{D^b}(\operatorname{mod}^{\mathbb{Z}} R) \to \underline{\mathrm{CM}}^{\mathbb{Z}} R,$$

where the first functor is induced from the surjective morphism $\Lambda \to \Lambda/(e)$, and the last functor is given by Theorem 2.10. Under Assumption 5.7(ii), F is shown to be a triangle equivalence. A crucial step is to show that F restricts to an equivalence $\mathrm{U}_d(\Lambda/(e)) \to \operatorname{add}\{\Gamma e(i) \mid i \in \mathbb{Z}\}$, which are d-cluster tilting subcategories of $\mathrm{D^b}(\operatorname{mod}\Lambda/(e))$ and $\underline{\mathrm{CM}}^{\mathbb{Z}} R$ respectively (Theorem 3.7). Similarly, we obtain a triangle equivalence $\mathrm{C}_d(\Lambda/(e)) \simeq \underline{\mathrm{CM}} R$ by using universality of d-cluster categories (Keller [2005]). As a summary, we obtain the following results.

Theorem 5.8 (Amiot, Iyama, and Reiten [2015]). *Under Assumption 5.7(i), let $R = e\Gamma e$ and $\Lambda = \Gamma_0$.*

(a) *R is a $(d+1)$-Iwanaga-Gorenstein algebra, and Γe is a CM R-module.*

(b) *Γe is a d-cluster tilting object in $\mathrm{CM}\,R$ and satisfies $\operatorname{End}_R(\Gamma e) = \Gamma$. Thus R is d-CM-finite and has a d-Auslander algebra Γ*

(c) *If Assumption 5.7(ii) is satisfied, then there exist triangle equivalences*

$$\mathrm{D^b}(\operatorname{mod}\Lambda/(e)) \simeq \underline{\mathrm{CM}}^{\mathbb{Z}} R \quad \text{and} \quad \mathrm{C}_d(\Lambda/(e)) \simeq \underline{\mathrm{CM}} R.$$

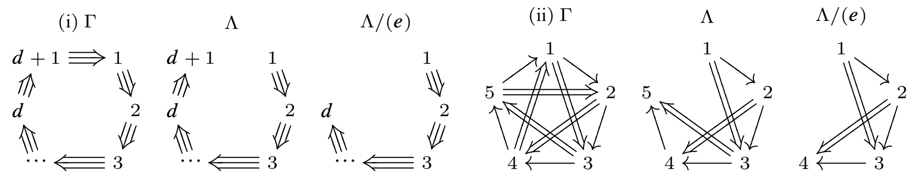
Similar triangle equivalences were given in de Völcsey and Van den Bergh [2016] and Kalck and Yang [2016] using different methods. There is a connection between (c) and (5-4) above via Theorem 2.12, see Amiot [2013].

In the case $d = 1$, the above (c) recovers Corollary 5.2 and a triangle equivalence $\underline{\mathrm{CM}} R \simeq \mathrm{C}_1(kQ/(e))$, which implies algebraic McKay correspondence in Example 2.25. Motivated by Example 3.14 and Theorem 5.1(c), we consider the following.

Example 5.9 (Amiot, Iyama, and Reiten [2015],Ueda [2008]). Let $S = k[x_0, \dots, x_d]$ be a polynomial algebra, and G a finite subgroup of $\mathrm{SL}_{d+1}(k)$. Then the skew group ring $\Gamma = S * G$ is a (ungraded) $(d+1)$-Calabi-Yau algebra. Assume that G is generated by the diagonal matrix $\operatorname{diag}(\zeta^{a_0}, \dots, \zeta^{a_d})$, where ζ is a primitive n-th root of unity and $0 \le a_j \le n-1$ for each j. Then Γ is presented by the McKay quiver of G, which has vertices $\mathbb{Z}/n\mathbb{Z}$, and arrows $x_j\colon i \to i + a_j$ for each i, j. Define a $\mathbb{Z}$-grading on Γ by $\deg(x_j\colon i \to i + a_j) = 0$ if $i < i + a_j$ as integers in $\{1, \dots, n\}$, and 1 otherwise. Then Γ is a $(d+1)$-Calabi-Yau algebra of a-invariant $-\sum_{0 \le j \le d} a_j/n$. Assume that this is -1, and let $e = e_n$. Then Assumption 5.7 is satisfied, and $e\Gamma e = S^G$ holds. Thus Theorem 5.8 gives triangle equivalences

$$\mathrm{D^b}(\operatorname{mod}\Lambda/(e)) \simeq \underline{\mathrm{CM}}^{\mathbb{Z}} S^G \quad \text{and} \quad \mathrm{C}_d(\Lambda/(e)) \simeq \underline{\mathrm{CM}} S^G.$$

Below we draw quivers for two cases (i) $n = d+1$ and $a_0 = \cdots = a_d = 1$, and (ii) $d = 2, n = 5$ and $(a_0, a_1, a_2) = (1, 2, 2)$.



In (i), $\Rrightarrow$ shows $d+1$ arrows, S^G is the Veronese subring $S^{(d+1)}$ and Λ is the Beilinson algebra. For $d = 2$, we recover the triangle equivalence $\mathrm{C}_2(kQ) \simeq \underline{\mathrm{CM}}S^G$ for $Q = [\,1 \Rrightarrow 2\,]$ given in Keller and Reiten [2008] and Keller, Murfet, and Van den Bergh [2011].

Note that similar triangle equivalences are given in Iyama and R. Takahashi [2013], Ueda [2012], and Mori and Ueyama [2016] for the skew group rings $S * G$ whose a-invariants are not equal to -1.

Example 5.10 (Dimer models)**.** Let G be a bipartite graph on a torus, and G_0 (resp. G_1, G_2) the set of vertices (resp. edges, faces) of G. We associate a quiver with potential (Q, W): The underlying graph of Q is the dual of the graph G, and faces of Q dual to white (resp. black) vertices are oriented clockwise (resp. anti-clockwise). Hence any vertex $v \in G_0$ corresponds to a cycle c_v of Q. Let $W = \sum_{v:\text{white}} c_v - \sum_{v:\text{black}} c_v$, and Γ the Jacobian algebra of (Q, W).

Under the assumption that G is *consistent*, Γ is a (ungraded) 3-Calabi-Yau algebra, and for any vertex e, $R = e\Gamma e$ is a Gorenstein toric singularity in dimension 3 (see Broomhead [2012] and Bocklandt [2012] and references therein). Using a perfect matching C on G, define a $\mathbb{Z}$-grading on Γ by $\deg a = 1$ for all $a \in C$ and $\deg a = 0$ otherwise. If both $\Gamma/(e)$ and $\Lambda = \Gamma_0$ are finite dimensional and $e\Lambda(1-e) = 0$ holds, then Theorem 5.8 gives triangle equivalences

$$\mathrm{D^b}(\mathrm{mod}\,\Lambda/(e)) \simeq \underline{\mathrm{CM}}^{\mathbb{Z}} R \quad\text{and}\quad \mathrm{C}_2(\Lambda/(e)) \simeq \underline{\mathrm{CM}} R.$$

5.3 d-representation-finite algebras.

In this subsection, we study the $(d+1)$-preprojective algebras of d-representation-finite algebras. We start with the following basic properties.

Proposition 5.11 (Geiss, Leclerc, and Schröer [2006], Iyama [2011], and Iyama and Oppermann [2013])**.** *Let Λ be a d-representation-finite k-algebra and $\Pi = \Pi_{d+1}(\Lambda)$.*

(a) *Π is a $\mathbb{Z}$-graded finite dimensional self-injective k-algebra.*

(b) *$\underline{\mathrm{mod}}^{\mathbb{Z}}\Pi$ has a Serre functor $(-1)[d+1]$, and $\underline{\mathrm{mod}}\Pi$ is $(d+1)$-Calabi-Yau.*

(c) Π *is a (unique) d-cluster tilting object in* $\operatorname{mod}\Lambda$.

Now we give an explicit characterization of such Π.

Definition 5.12. Let $\Gamma = \bigoplus_{i\geq 0}\Gamma_i$ be a $\mathbb{Z}$-graded finite dimensional self-injective k-algebra. We denote by $\Gamma^{\mathrm{e}} = \Gamma^{\mathrm{op}}\otimes_k\Gamma$ the enveloping algebra of Γ. We say that Γ is a *stably d-Calabi-Yau algebra of a-invariant a* (or *Gorenstein parameter $-a$*) if $\mathbf{R}\mathrm{Hom}_{\Gamma^{\mathrm{e}}}(\Gamma,\Gamma^{\mathrm{e}})(a)[d] \simeq \Gamma$ in $\mathrm{D}^{\mathbb{Z}}_{\mathrm{sg}}(\Gamma^{\mathrm{e}})$.

Now we give a homological characterization of the $(d+1)$-preprojective algebras of d-representation-finite algebras as the explicit correspondence.

Theorem 5.13 (Amiot and Oppermann [2014])**.** *There exists a bijection between the set of isomorphism classes of d-representation-finite algebras Λ and the set of isomorphism classes of stably $(d+1)$-Calabi-Yau self-injective algebras Γ of a-invariant -1. It is given by $\Lambda\mapsto\Pi_{d+1}(\Lambda)$ and $\Gamma\mapsto\Gamma_0$.*

Now let Λ be a d-representation-finite k-algebra, and $\Pi = \Pi_{d+1}(\Lambda)$. Let $\Gamma = \underline{\mathrm{End}}_{\Lambda}(\Pi)$ be the stable d-Auslander algebra of Λ. Then we have an equivalence

$$\mathrm{U}_d(\Lambda) \simeq \mathrm{proj}^{\mathbb{Z}}\, T(\Gamma) \tag{5-5}$$

of additive categories. Thus we have triangle equivalences

$$\underline{\mathrm{mod}}^{\mathbb{Z}}\Pi \overset{(5\text{-}3)}{\simeq} \underline{\mathrm{mod}}\mathrm{U}_d(\Lambda) \overset{(5\text{-}5)}{\simeq} \underline{\mathrm{mod}}^{\mathbb{Z}}T(\Gamma) \overset{(4\text{-}1)}{\simeq} \mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Gamma).$$

By Proposition 5.11(b), the automorphism (-1) on $\underline{\mathrm{mod}}^{\mathbb{Z}}\Pi$ corresponds to ν_{d+1} on $\mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Gamma)$. Using universality of $(d+1)$-cluster categories (Keller [2005]), we obtain a triangle equivalence $\underline{\mathrm{mod}}\Pi \simeq \mathrm{C}_{d+1}(\Gamma)$. As a summary, we obtain the following.

Theorem 5.14 (Iyama and Oppermann [2013])**.** *Let Λ be a d-representation-finite k-algebra, $\Pi = \Pi_{d+1}(\Lambda)$, and $\Gamma = \underline{\mathrm{End}}_{\Lambda}(\Pi)$ the stable d-Auslander algebra of Λ. Then there exist triangle equivalences*

$$\underline{\mathrm{mod}}^{\mathbb{Z}}\Pi \simeq \mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Gamma) \ \textit{and}\ \underline{\mathrm{mod}}\Pi \simeq \mathrm{C}_{d+1}(\Gamma).$$

Applying Theorem 5.14 for $d = 1$, we obtain the following observations.

Example 5.15 (Amiot [2009] and Iyama and Oppermann [2013])**.** Let Π be the preprojective algebra of a Dynkin quiver Q, and Γ the stable Auslander algebra of kQ. Then there exist triangle equivalences

$$\underline{\mathrm{mod}}^{\mathbb{Z}}\Pi \simeq \mathrm{D}^{\mathrm{b}}(\operatorname{mod}\Gamma) \ \text{and}\ \underline{\mathrm{mod}}\Pi \simeq \mathrm{C}_2(\Gamma).$$

As an application, if a quiver Q' has the same underlying graph with Q, then the stable Auslander algebra Γ' of kQ' is derived equivalent to Γ since Π is common.

In the rest of this subsection, we discuss properties of $\Pi_{d+1}(\Lambda)$ for a more general class of Λ. We say that a finite dimensional k-algebra Λ with $\operatorname{gl.dim}\Lambda \le d$ satisfies the *vosnex property* if Λ is ν_d-finite and satisfies $\operatorname{Hom}_{\mathrm{D^b}(\operatorname{mod}\Lambda)}(\mathrm{U}_d(\Lambda)[i], \mathrm{U}_d(\Lambda)) = 0$ for all $1 \le i \le d-2$. This is automatic if $d = 1, 2$ or Λ is d-representation-finite. In this case, the following generalization of Theorem 5.14 holds.

Theorem 5.16 (Iyama and Oppermann [2013]). *Let Λ be a finite dimensional k-algebra with* $\operatorname{gl.dim}\Lambda \le d$ *satisfying the vosnex property. Then $\Pi = \Pi_{d+1}(\Lambda)$ is 1-Iwanaga-Gorenstein, $\Gamma = \underline{\operatorname{End}}_\Lambda(\Pi)$ satisfies* $\operatorname{gl.dim}\Gamma \le d+1$*, and there exist triangle equivalences*

$$\underline{\mathrm{CM}}^{\mathbb{Z}}\Pi \simeq \mathrm{D^b}(\operatorname{mod}\Gamma) \ \textit{and}\ \underline{\mathrm{CM}}\Pi \simeq \mathrm{C}_{d+1}(\Gamma).$$

For more general Λ, we refer to Beligiannis [2015] for some properties of $\Pi_{d+1}(\Lambda)$.

5.4 Preprojective algebras and Coxeter groups. We discuss a family of finite dimensional k-algebras constructed from preprojective algebras and Coxeter groups.

Let Q be an acyclic quiver and Π the preprojective algebra of kQ. The *Coxeter group* of Q is generated by s_i with $i \in Q_0$, and the relations are the following.

- $s_i^2 = 1$ for all $i \in Q_0$.
- $s_i s_j = s_j s_i$ if there is no arrow between i and j in Q.
- $s_i s_j s_i = s_j s_i s_j$ if there is precisely one arrow between i and j in Q.

Let $w \in W$. An expression $w = s_{i_1} s_{i_2} \cdots s_{i_\ell}$ of w is called *reduced* if ℓ is minimal among all expressions of w. For $i \in Q_0$, let I_i be the two-sided ideal of Π generated by the idempotent $1 - e_i$. For a reduced expression $w = s_{i_1} \cdots s_{i_\ell}$, we define a two-sided ideal of Π by

$$I_w := I_{i_1} I_{i_2} \cdots I_{i_\ell}.$$

This is independent of the choice of the reduced expression of w. The corresponding factor algebra $\Pi_w := \Pi/I_w$ is a finite dimensional k-algebra. It enjoys the following remarkable properties.

Theorem 5.17 (Buan, Iyama, Reiten, and Scott [2009], Geiss, Leclerc, and Schröer [2007], and Amiot, Reiten, and Todorov [2011]). *Let $w \in W$.*

(a) *Π_w is a 1-Iwanaga-Gorenstein algebra.*

(b) *$\underline{\mathrm{CM}}\Pi_w$ is a 2-Calabi-Yau triangulated category.*

(c) *There exists a 2-cluster tilting object $\bigoplus_{j=1}^{\ell} e_{i_j}\Pi_{s_{i_j}\cdots s_{i_\ell}}$ in $\mathrm{CM}\,\Pi_w$.*

(d) *There exists a triangle equivalence* $\underline{\mathrm{CM}}\Pi_w \simeq \mathrm{C}_2(\Lambda)$ *for some algebra* Λ.

Therefore it is natural to expect that there exists a triangle equivalence $\underline{\mathrm{CM}}^{\mathbb{Z}}\Pi_w \simeq \mathrm{D}^{\mathrm{b}}(\mathrm{mod}\,\Lambda')$ for some algebra Λ'. In fact, the following results are known, where we refer to Kimura [2018, 2016] for the definitions of *c-sortable*, *c-starting* and *c-ending*.

Theorem 5.18. *Let* $w = s_{i_1} \cdots s_{i_\ell}$ *be a reduced expression of* $w \in W$.

(a) *(Kimura [2018]) If w is c-sortable, then* $\underline{\mathrm{CM}}^{\mathbb{Z}}\Pi_w$ *has a tilting object* $\bigoplus_{i>0} \Pi_w(i)_{\geq 0}$.

(b) *(Kimura [2016])* $\underline{\mathrm{CM}}^{\mathbb{Z}}\Pi_w$ *has a silting object* $\bigoplus_{j=1}^{\ell} e_{i_j}\Pi_{s_{i_j}\cdots s_{i_\ell}}$. *This is a tilting object if the reduced expression is c-starting or c-ending.*

We end this section with posing the following natural question on 'higher cluster combinatorics' (e.g. Oppermann and Thomas [2012]), which will be related to derived equivalences of Calabi-Yau algebras since our I_w is a tilting object in $\mathrm{K}^{\mathrm{b}}(\mathrm{proj}\,\Pi)$ if Q is non-Dynkin.

Problem 5.19. *Are there similar results to Theorems 5.17 and 5.18 for higher preprojective algebras? What kind of combinatorial structure will appear instead of the Coxeter groups?*

6 Geigle-Lenzing complete intersections

Weighted projective lines of Geigle and Lenzing [1987] are one of the basic objects in representation theory. For example, the simplest class of weighted projective lines gives us simple singularities in dimension 2 as certain Veronese subrings. We introduce a higher dimensional generalization of weighted projective lines following Herschend, Iyama, Minamoto, and Oppermann [2014].

6.1 Basic properties. For a field k and an integer $d \geq 1$, we consider a polynomial algebra $C = k[T_0, \dots, T_d]$. For $n \geq 0$, let $\ell_1, \dots, \ell_n$ be linear forms in C and $p_1, \dots, p_n$ positive integers. For simplicity, we assume $p_i \geq 2$ for all i. Let

$$R = C[X_1, \dots, X_n]/(X_i^{p_i} - \ell_i \mid 1 \leq i \leq n)$$

be the factor algebra of the polynomial algebra $C[X_1, \dots, X_n]$, and

$$\mathbb{L} = \langle \vec{x}_1, \dots, \vec{x}_n, \vec{c}\rangle / \langle p_i\vec{x}_i - \vec{c} \mid 1 \leq i \leq n\rangle.$$

the factor group of the free abelian group $\langle \vec{x}_1, \dots, \vec{x}_n, \vec{c}\rangle$. Then $\mathbb{L}$ is an abelian group of rank 1 with torsion elements in general, and R is $\mathbb{L}$-graded by $\deg T_i = \vec{c}$ for all $0 \leq i \leq d$ and $\deg X_i = \vec{x}_i$ for all $1 \leq i \leq n$.

We call the pair $(R, \mathbb{L})$ a *Geigle-Lenzing* (*GL*) *complete intersection* if $\ell_1, \dots, \ell_n$ are in general position in the sense that each set of at most $d+1$ elements from $\ell_1, \dots, \ell_n$ is linearly independent. We give some basic properties.

Proposition 6.1. *Let $(R, \mathbb{L})$ be a GL complete intersection.*

(a) *$X_1^{p_1} - \ell_1, \dots, X_n^{p_n} - \ell_n$ is a $C[X_1, \dots, X_n]$-regular sequence.*

(b) *R is a complete intersection ring with $\dim R = d+1$ and has an a-invariant*

$$\vec{\omega} = (n-d-1)\vec{c} - \sum_{i=1}^{n} \vec{x}_i .$$

(c) *After a suitable linear transformation of variables $T_0, \dots, T_d$, we have*

$$R = \begin{cases} k[X_1, \dots, X_n, T_n, \dots, T_d] & \text{if } n \le d+1, \\ k[X_1, \dots, X_n]/(X_i^{p_i} - \sum_{j=1}^{d+1} \lambda_{i,j-1} X_j^{p_j}) \mid d+2 \le i \le n) & \text{if } n \ge d+2. \end{cases}$$

(d) *R is regular if and only if R is a polynomial algebra if and only if $n \le d+1$.*

(e) *$\mathrm{CM}^{\mathbb{L}} R = \mathrm{CM}_0^{\mathbb{L}} R$ holds, and $\underline{\mathrm{CM}}^{\mathbb{L}} R$ has a Serre functor $(\vec{\omega})[d]$ (Theorem 2.16).*

Let $\delta\colon \mathbb{L} \to \mathbb{Q}$ be a group homomorphism given by $\delta(\vec{x}_i) = \frac{1}{p_i}$ and $\delta(\vec{c}) = 1$. We consider the following trichotomy given by the sign of $\delta(\vec{\omega}) = n - d - 1 - \sum_{i=1}^{n} \frac{1}{p_i}$. For example, $(R, \mathbb{L})$ is Fano if $n \le d+1$.

$\delta(\vec{\omega})$	< 0	$= 0$	> 0
$(R, \mathbb{L})$	Fano	Calabi-Yau	anti-Fano
$d = 1$	domestic	tubular	wild

In the classical case $d = 1$, the ring R has been studied in the context of weighted projective lines. The above trichotomy is given explicitly as follows.

- 5 types for domestic: $n \le 2$, $(2, 2, p)$, $(2, 3, 3)$, $(2, 3, 4)$ and $(2, 3, 5)$.
- 4 types for tubular: $(3, 3, 3)$, $(2, 4, 4)$, $(2, 3, 6)$ and $(2, 2, 2, 2)$.
- All other types are wild.

There is a close connection between domestic type and simple singularities. The following explains Corollary 5.2, where $R^{(\vec{\omega})} = \bigoplus_{i \in \mathbb{Z}} R_{i\vec{\omega}}$ is the Veronese subring.

Theorem 6.2 (Geigle and Lenzing [1991]). *If $(R, \mathbb{L})$ is domestic, then $R^{(\vec{\omega})}$ is a simple singularity $k[x, y, z]/(f)$ in dimension 2, and we have an equivalence $\mathrm{CM}^{\mathbb{L}} R \simeq \mathrm{CM}^{\mathbb{Z}} R^{(\vec{\omega})}$. The AR quiver is $\mathbb{Z}Q$, where Q is given by the following table.*

$(p_1, \dots, p_n)$	x	y	z	f	Q
(p, q)	X_1X_2	X_2^{p+q}	X_1^{p+q}	$x^{p+q} - yz$	$\widetilde{\mathbb{A}}_{p,q}$
$(2, 2, 2p)$	X_3^2	X_1^2	$X_1X_2X_3$	$x(y^2 + x^p y) + z^2$	$\widetilde{\mathbb{D}}_{2p+2}$
$(2, 2, 2p+1)$	X_3^2	X_1X_2	$X_1^2X_3$	$x(y^2 + x^p z) + z^2$	$\widetilde{\mathbb{D}}_{2p+3}$
$(2, 3, 3)$	X_1	X_2X_3	X_2^3	$x^2z + y^3 + z^2$	$\widetilde{\mathbb{E}}_6$
$(2, 3, 4)$	X_2	X_3^2	X_1X_3	$x^3y + y^3 + z^2$	$\widetilde{\mathbb{E}}_7$
$(2, 3, 5)$	X_3	X_2	X_1	$x^5 + y^3 + z^2$	$\widetilde{\mathbb{E}}_8$

6.2 Cohen-Macaulay representations. To study the category $\mathrm{CM}^{\mathbb{L}} R$, certain finite dimensional algebras play an important role. For a finite subset I of $\mathbb{L}$, let

$$A^I = \bigoplus_{\vec{x},\vec{y} \in I} R_{\vec{x}-\vec{y}}.$$

We define the multiplication in A^I by $(r_{\vec{x},\vec{y}})_{\vec{x},\vec{y}\in I} \cdot (r'_{\vec{x},\vec{y}})_{\vec{x},\vec{y}\in I} = (\sum_{\vec{z}\in I} r_{\vec{x},\vec{z}} r'_{\vec{z},\vec{y}})_{\vec{x},\vec{y}\in I}$. Then A^I forms a finite dimensional k-algebra called the *I-canonical algebra*.

We define a partial order $\leq$ on $\mathbb{L}$ by writing $\vec{x} \leq \vec{y}$ if $\vec{y} - \vec{x}$ belongs to $\mathbb{L}_+$, where $\mathbb{L}_+$ is the submonoid of $\mathbb{L}$ generated by $\vec{c}$ and $\vec{x}_i$ for all i. For $\vec{x} \in \mathbb{L}$, let $[0, \vec{x}]$ be the interval in $\mathbb{L}$, and $A^{[0,\vec{x}]}$ the $[0, \vec{x}]$-canonical algebra. We call

$$A^{\mathrm{CM}} = A^{[0, d\vec{c}+2\vec{\omega}]}$$

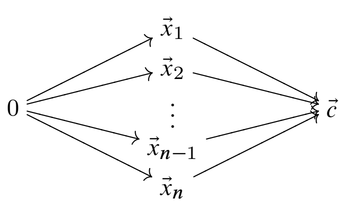
the *CM-canonical algebra*.

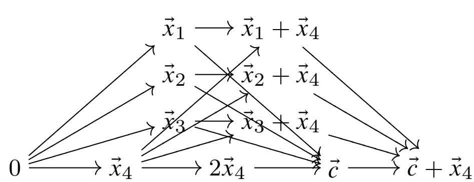
Example 6.3. The equality $d\vec{c} + 2\vec{\omega} = (n - d - 2)\vec{c} + \sum_{i=1}^{n} (p_i - 2)\vec{x}_i$ holds.

(a) If $n \leq d + 1$, then $A^{\mathrm{CM}} = 0$. If $n = d + 2$, then $A^{\mathrm{CM}} = \bigotimes_{i=1}^{n+2} k\mathbb{A}_{p_i-1}$.

(b) If $n = d + 3$ and $p_i = 2$ for all i, then A^{CM} has the left quiver below.

(c) If $d = 1$, $n = 4$ and $(p_i)_{i=1}^4 = (2, 2, 2, 3)$, then A^{CM} has the right quiver below.





The following is a main result in this section.

Theorem 6.4. *Let $(R, \mathbb{L})$ be a GL complete intersection. Then there is a triangle equivalence*

$$\underline{\mathrm{CM}}^{\mathbb{L}} R \simeq \mathrm{D}^{\mathrm{b}}(\mathrm{mod}\, A^{\mathrm{CM}}).$$

In particular, $\underline{\mathrm{CM}}^{\mathbb{L}} R$ has a tilting object.

The case $n = d + 2$ was shown in Kussin, Lenzing, and Meltzer [2013] ($d = 1$) and Futaki and Ueda [2011]. An important tool in the proof is an $\mathbb{L}$-analogue of Theorem 2.12.

As an application, one can immediately obtain the following analogue of Theorem 2.23 by using the knowledge on A^{CM} in representation theory, where we call $(R, \mathbb{L})$ *CM-finite* if there are only finitely many isomorphism classes of indecomposable objects in $\mathrm{CM}^{\mathbb{L}} R$ up to degree shift (cf. Definition 2.22).

Corollary 6.5. *Let $(R, \mathbb{L})$ be a GL complete intersection. Then $(R, \mathbb{L})$ is CM-finite if and only if one of the following conditions hold.*

(i) $n \leq d + 1$.

(ii) $n = d + 2$, *and* $(p_1, \dots, p_n) = (2, \dots, 2, p_n)$, $(2, \dots, 2, 3, 3)$, $(2, \dots, 2, 3, 4)$ *or* $(2, \dots, 2, 3, 5)$ *up to permutation.*

We call a GL complete intersection $(R, \mathbb{L})$ *d-CM-finite* if there exists a d-cluster tilting subcategory $\mathcal{C}$ of $\mathrm{CM}^{\mathbb{L}} R$ such that there are only finitely many isomorphism classes of indecomposable objects in $\mathcal{C}$ up to degree shift (cf. Definition 3.12). Now we discuss which GL complete intersections are d-CM-finite. Our Theorem 3.7 gives the following sufficient condition, where a tilting object is called *d-tilting* if the endomorphism algebra has global dimension at most d.

Proposition 6.6. *If $\underline{\mathrm{CM}}^{\mathbb{L}} R$ has a d-tilting object U, then $(R, \mathbb{L})$ is d-CM-finite and $\mathrm{CM}^{\mathbb{L}} R$ has the d-cluster tilting subcategory* $\mathrm{add}\{U(\ell\vec{\omega}), R(\vec{x}) \mid \ell \in \mathbb{Z}, \vec{x} \in \mathbb{L}\}$.

Therefore the following problem is of our interest.

Problem 6.7. *When does $\underline{\mathrm{CM}}^{\mathbb{L}} R$ have a d-tilting object? Equivalently, when is A^{CM} derived equivalent to an algebra Λ with* $\mathrm{gl.dim}\, \Lambda \leq d$?

Applying Tate's DG algebra resolutions (Tate [1957]), we can calculate $\mathrm{gl.dim}\, A^{\mathrm{CM}}$. Note that any element $\vec{x} \in \mathbb{L}$ can be written uniquely as $\vec{x} = a\vec{c} + \sum_{i=1}^{n} a_i \vec{x}_i$ for $a \in \mathbb{Z}$ and $0 \leq a_i \leq p_i - 1$, which is called the *normal form* of $\vec{x}$.

Theorem 6.8. (a) *Write $\vec{x} \in \mathbb{L}_+$ in normal form $\vec{x} = a\vec{c} + \sum_{i=1}^{n} a_i \vec{x}_i$. Then*

$$\mathrm{gl.dim}\, A^{[0,\vec{x}]} = \begin{cases} \min\{d + 1, a + \#\{i \mid a_i \neq 0\}\} & \text{if } n \leq d + 1, \\ 2a + \#\{i \mid a_i \neq 0\} & \text{if } n \geq d + 2. \end{cases}$$

(b) *If* $n \geq d+2$, *then* A^{CM} *has global dimension* $2(n-d-2)+\#\{i \mid p_i \geq 3\}$.

We obtain the following examples from Theorem 6.8 and the fact that $k\mathbb{A}_2 \otimes_k k\mathbb{A}_m$ is derived equivalent to $k\mathbb{D}_4$ if $m=2$, $k\mathbb{E}_6$ if $m=3$, and $k\mathbb{E}_8$ if $m=4$.

Example 6.9. In the following cases, $\underline{\mathrm{CM}}^{\mathbb{L}} R$ has a d-tilting object.

(i) $n \leq d+1$.

(ii) $n = d+2 \geq 3$ and $(p_1, p_2, p_3) = (2,2,p_3)$, $(2,3,3)$, $(2,3,4)$ or $(2,3,5)$.

(iii) $n = d+2 \geq 4$ and $(p_1,p_2,p_3,p_4) = (3,3,p_3,p_4)$ with $p_3, p_4 \in \{3,4,5\}$.

(iv) $\#\{i \mid p_i = 2\} \geq 3(n-d)-4$.

The following gives a necessary condition for the existence of d-tilting object.

Proposition 6.10. *If* $\underline{\mathrm{CM}}^{\mathbb{L}} R$ *has a* d*-tilting object, then* $(R, \mathbb{L})$ *is Fano.*

Note that the converse is not true. For example, let $d=2$ and $(2,5,5,5)$. Then $(R,\mathbb{L})$ is Fano since $\delta(\vec{\omega}) = -\frac{1}{10}$. On the other hand, $A^{\mathrm{CM}} = \bigotimes_{i=1}^{3} k\mathbb{A}_4$ satisfies $\nu^5 = [9]$. One can show that A^{CM} is not derived equivalent to an algebra Λ with $\mathrm{gl.dim}\,\Lambda \leq 2$ by using the inequality $2(5-1) < 9$.

6.3 Geigle-Lenzing projective spaces. Let $(R,\mathbb{L})$ be a GL complete intersection. Recall that $\mathrm{mod}_0^{\mathbb{L}} R$ is the Serre subcategory of $\mathrm{mod}^{\mathbb{L}} R$ consisting of finite dimensional modules. We consider the quotient category

$$\mathrm{coh}\,\mathbb{X} = \mathrm{qgr}\,R = \mathrm{mod}^{\mathbb{L}} R / \mathrm{mod}_0^{\mathbb{L}} R.$$

We call objects in $\mathrm{coh}\,\mathbb{X}$ *coherent sheaves* on the *GL projective space* $\mathbb{X}$. We can regard $\mathbb{X}$ as the quotient stack $[(\mathrm{Spec}\,R \setminus \{R_+\}) / \mathrm{Spec}\,k[\mathbb{L}]]$ for $R_+ = \bigoplus_{\vec{x}>0} R_{\vec{x}}$. For example, if $n=0$, then $\mathbb{X}$ is the projective space $\mathbb{P}^d$.

We study the bounded derived category $\mathrm{D^b}(\mathrm{coh}\,\mathbb{X})$, which is canonically triangle equivalent to the Verdier quotient $\mathrm{D^b}(\mathrm{mod}^{\mathbb{L}} R)/\mathrm{D^b}(\mathrm{mod}_0^{\mathbb{L}} R)$. The duality $(-)^* = \mathbf{R}\mathrm{Hom}_R(-,R)\colon \mathrm{D^b}($ $\mathrm{D^b}(\mathrm{mod}^{\mathbb{L}} R)$ induces a duality $(-)^*\colon \mathrm{D^b}(\mathrm{coh}\,\mathbb{X}) \to \mathrm{D^b}(\mathrm{coh}\,\mathbb{X})$. We define the category of *vector bundles* on $\mathbb{X}$ as

$$\mathrm{vect}\,\mathbb{X} = \mathrm{coh}\,\mathbb{X} \cap (\mathrm{coh}\,\mathbb{X})^*.$$

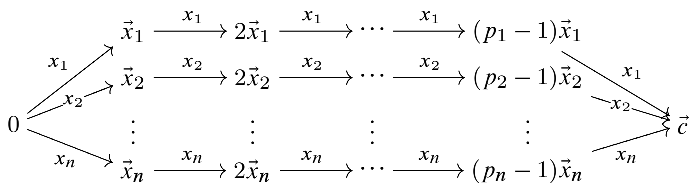
The composition $\mathrm{CM}^{\mathbb{L}} R \subset \mathrm{mod}^{\mathbb{L}} R \to \mathrm{coh}\,\mathbb{X}$ is fully faithful, and we can regard $\mathrm{CM}^{\mathbb{L}} R$ as a full subcategory of $\mathrm{vect}\,\mathbb{X}$. We have $\mathrm{CM}^{\mathbb{L}} R = \mathrm{vect}\,\mathbb{X}$ if $d=1$, but this is not the case if $d \geq 2$. In fact, we have equalities

$$\text{(6-1)} \qquad \mathrm{CM}^{\mathbb{L}} R = \{X \in \operatorname{vect} \mathbb{X} \mid \forall \vec{x} \in \mathbb{L},\ 1 \leq i \leq d-1,\ \operatorname{Ext}^i_{\mathbb{X}}(\mathcal{O}(\vec{x}), X) = 0\}$$
$$= \{X \in \operatorname{vect} \mathbb{X} \mid \forall \vec{x} \in \mathbb{L},\ 1 \leq i \leq d-1,\ \operatorname{Ext}^i_{\mathbb{X}}(X, \mathcal{O}(\vec{x})) = 0\}$$

where $\mathcal{O}(\vec{x}) = R(\vec{x})$. Now we define the *$d$-canonical algebra* by

$$A^{\mathrm{ca}} = A^{[0, d\vec{c}]}.$$

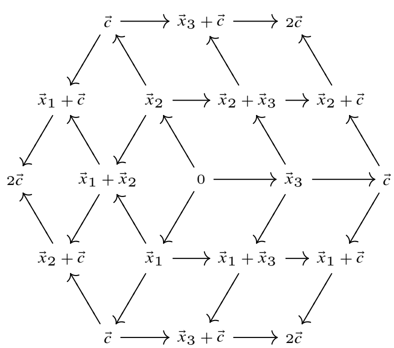
Example 6.11. (a) If $d = 1$, then A^{ca} is precisely the canonical algebra of Ringel [1984]. It is given by the following quiver with relations $x_i^{p_i} = \lambda_{i0} x_1^{p_1} + \lambda_{i1} x_2^{p_2}$ for any i with $3 \leq i \leq n$.

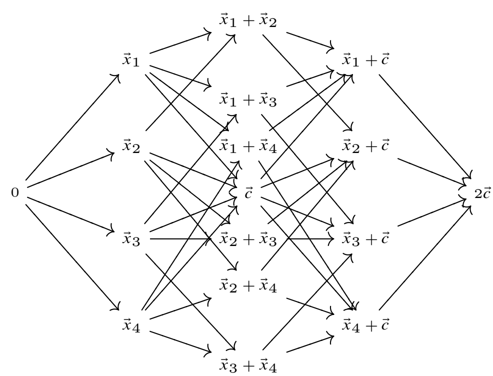


(b) If $n = 0$, then A^{ca} is the Beilinson algebra.

(c) If $d = 2$, $n = 3$ and $(p_i)_{i=1}^3 = (2, 2, 2)$, then A^{ca} has the left quiver below.

(d) If $d = 2$, $n = 4$ and $(p_i)_{i=1}^4 = (2, 2, 2, 2)$, then A^{ca} has the right quiver below.





As in the case of $\underline{\mathrm{CM}}^{\mathbb{L}} R$ and A^{CM}, we obtain the following results.

Theorem 6.12. *Let $\mathbb{X}$ be a GL projective space. Then there is a triangle equivalence*

$$\mathrm{D^b}(\operatorname{coh} \mathbb{X}) \simeq \mathrm{D^b}(\operatorname{mod} A^{\mathrm{ca}}).$$

Moreover $\mathrm{D^b}(\operatorname{coh} \mathbb{X})$ *has a tilting bundle* $\bigoplus_{\vec{x} \in [0, d\vec{c}]} \mathcal{O}(\vec{x})$.

Some cases were known before ($n = 0$ by Beilinson [1978], $d = 1$ by Geigle and Lenzing [1987], $n \leq d+1$ by Baer [1988], $n = d+2$ by Ishii and Ueda [2012]). An important tool in the proof is again an $\mathbb{L}$-analogue of Theorem 2.12.

We call $\mathbb{X}$ *vector bundle finite* (*VB-finite*) if there are only finitely many isomorphism classes of indecomposable objects in $\operatorname{vect}\mathbb{X}$ up to degree shift. There is a complete classification: $\mathbb{X}$ is VB-finite if and only if $d = 1$ and $\mathbb{X}$ is domestic.

We call $\mathbb{X}$ *d-VB-finite* if there exists a d-cluster tilting subcategory $\mathcal{C}$ of $\operatorname{vect}\mathbb{X}$ such that there are only finitely many isomorphism classes of indecomposable objects in $\mathcal{C}$ up to degree shift. In the rest, we discuss which GL projective spaces are d-VB-finite. We start with the following relationship between d-cluster tilting subcategories of $\mathrm{CM}^{\mathbb{L}} R$ and $\operatorname{vect}\mathbb{X}$, which follows from (6-1).

Proposition 6.13. *The d-cluster-tilting subcategories of $\mathrm{CM}^{\mathbb{L}} R$ are precisely the d-cluster-tilting subcategories of $\operatorname{vect}\mathbb{X}$ containing $\mathcal{O}(\vec{x})$ for all $\vec{x} \in \mathbb{L}$. Therefore, if $(R, \mathbb{L})$ is d-CM-finite, then $\mathbb{X}$ is d-VB-finite.*

For example, if $n \leq d+1$, then $\mathrm{CM}^{\mathbb{L}} R = \operatorname{proj}^{\mathbb{L}} R$ is a d-cluster tilting subcategory of itself, and hence $\operatorname{vect}\mathbb{X}$ has a d-cluster tilting subcategory $\operatorname{add}\{\mathcal{O}(\vec{x}) \mid \vec{x} \in \mathbb{L}\}$. This implies Horrocks' splitting criterion for $\operatorname{vect}\mathbb{P}^d$ (Okonek, Schneider, and Spindler [1980]).

We give another sufficient condition for d-VB-finiteness. Recall that we call a tilting object V in $\mathrm{D}^{\mathrm{b}}(\operatorname{coh}\mathbb{X})$ *d-tilting* if $\operatorname{gl.dim}\operatorname{End}_{\mathrm{D}^{\mathrm{b}}(\operatorname{coh}\mathbb{X})}(V) \leq d$.

Proposition 6.14. *Let $\mathbb{X}$ be a GL projective space, and V a d-tilting object in $\mathrm{D}^{\mathrm{b}}(\operatorname{coh}\mathbb{X})$.*

(a) *(cf. Example 3.11(b))* $\operatorname{gl.dim}\operatorname{End}_{\mathrm{D}^{\mathrm{b}}(\operatorname{coh}\mathbb{X})}(V) = d$ *holds. If $V \in \operatorname{coh}\mathbb{X}$, then $\operatorname{End}_{\mathbb{X}}(T)$ is a d-representation-infinite algebra.*

(b) *If $V \in \operatorname{vect}\mathbb{X}$, then $\mathbb{X}$ is d-VB-finite and $\operatorname{vect}\mathbb{X}$ has the d-cluster tilting subcategory* $\operatorname{add}\{V(\ell\vec{\omega}) \mid \ell \in \mathbb{Z}\}$.

Therefore it is natural to ask when $\mathbb{X}$ has a d-tilting bundle, or equivalently, when A^{ca} is derived equivalent to an algebra Λ with $\operatorname{gl.dim}\Lambda = d$. It follows from Theorem 6.8(a) that

$$\operatorname{gl.dim} A^{\mathrm{ca}} = \begin{cases} d & \text{if } n \leq d+1, \\ 2d & \text{if } n \geq d+2. \end{cases}$$

Thus, if $n \leq d+1$, then $\mathbb{X}$ has a d-tilting bundle. Using Example 6.9 and some general results on matrix factorizations, we have more examples.

Theorem 6.15. *In the following cases, $\mathbb{X}$ has a d-tilting bundle.*

(i) $n \leq d+1$.

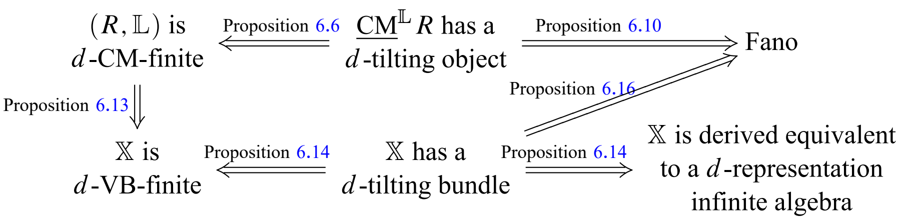
(ii) $n = d + 2 \geq 3$ *and* $(p_1, p_2, p_3) = (2, 2, p_3)$, $(2, 3, 3)$, $(2, 3, 4)$ *or* $(2, 3, 5)$.

(iii) $n = d + 2 \geq 4$ *and* $(p_1, p_2, p_3, p_4) = (3, 3, p_3, p_4)$ *with* $p_3, p_4 \in \{3, 4, 5\}$.

As in the previous subsection, we have the following necessary condition.

Proposition 6.16. *If* $\mathbb{X}$ *has a* d*-tilting bundle, then* $\mathbb{X}$ *is Fano.*

Some of our results in this section can be summarized as follows.



It is important to understand the precise relationship between these conditions. We refer to Chan [2017] and Buchweitz, Hille, and Iyama [n.d.] for results on existence of d-tilting bundles on more general varieties and stacks.

References

Takuma Aihara and Osamu Iyama (2012). "Silting mutation in triangulated categories". *J. Lond. Math. Soc. (2)* 85.3, pp. 633–668. MR: 2927802 (cit. on p. 154).

Claire Amiot (2009). "Cluster categories for algebras of global dimension 2 and quivers with potential". *Ann. Inst. Fourier (Grenoble)* 59.6, pp. 2525–2590. MR: 2640929 (cit. on pp. 155, 156, 166).

– (2013). "Preprojective algebras, singularity categories and orthogonal decompositions". In: *Algebras, quivers and representations*. Vol. 8. Abel Symp. Springer, Heidelberg, pp. 1–11. MR: 3183878 (cit. on p. 164).

Claire Amiot, Osamu Iyama, and Idun Reiten (2015). "Stable categories of Cohen-Macaulay modules and cluster categories". *Amer. J. Math.* 137.3, pp. 813–857. MR: 3357123 (cit. on pp. 144, 163, 164).

Claire Amiot and Steffen Oppermann (2014). "Higher preprojective algebras and stably Calabi-Yau properties". *Math. Res. Lett.* 21.4, pp. 617–647. MR: 3275638 (cit. on p. 166).

Claire Amiot, Idun Reiten, and Gordana Todorov (2011). "The ubiquity of generalized cluster categories". *Adv. Math.* 226.4, pp. 3813–3849. MR: 2764906 (cit. on p. 167).

Lidia Angeleri Hügel, Dieter Happel, and Henning Krause, eds. (2007). *Handbook of tilting theory*. Vol. 332. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, pp. viii+472. MR: 2385175 (cit. on p. 154).

Tokuji Araya (1999). "Exceptional sequences over graded Cohen-Macaulay rings". *Math. J. Okayama Univ.* 41, 81–102 (2001). MR: 1816620 (cit. on p. 160).

Michael Artin and William F. Schelter (1987). "Graded algebras of global dimension 3". *Adv. in Math.* 66.2, pp. 171–216. MR: 917738 (cit. on p. 146).

Michael Artin and Jean-Louis Verdier (1985). "Reflexive modules over rational double points". *Math. Ann.* 270.1, pp. 79–82. MR: 769609 (cit. on p. 153).

Michael Artin and James J. Zhang (1994). "Noncommutative projective schemes". *Adv. in Math.* 109.2, pp. 228–287 (cit. on p. 149).

Ibrahim Assem, Daniel Simson, and Andrzej Skowroński (2006). *Elements of the representation theory of associative algebras. Vol. 1*. Vol. 65. London Mathematical Society Student Texts. Techniques of representation theory. Cambridge University Press, Cambridge, pp. x+458 (cit. on pp. 146, 151).

Maurice Auslander (1978). "Functors and morphisms determined by objects", 1–244. Lecture Notes in Pure Appl. Math., Vol. 37. MR: 0480688 (cit. on pp. 150, 161).

– (1986). "Rational singularities and almost split sequences". *Trans. Amer. Math. Soc.* 293.2, pp. 511–531. MR: 816307 (cit. on p. 153).

Maurice Auslander and Mark Bridger (1969). *Stable module theory*. Memoirs of the American Mathematical Society, No. 94. American Mathematical Society, Providence, R.I., p. 146. MR: 0269685 (cit. on p. 147).

Maurice Auslander and Ragnar-Olaf Buchweitz (1989). "The homological theory of maximal Cohen-Macaulay approximations". *Mém. Soc. Math. France (N.S.)* 38. Colloque en l'honneur de Pierre Samuel (Orsay, 1987), pp. 5–37. MR: 1044344 (cit. on p. 144).

Maurice Auslander and Idun Reiten (1987). "Almost split sequences for $\mathbf{Z}$-graded rings". In: *Singularities, representation of algebras, and vector bundles (Lambrecht, 1985)*. Vol. 1273. Lecture Notes in Math. Springer, Berlin, pp. 232–243. MR: 915178 (cit. on p. 150).

– (1991). "Applications of contravariantly finite subcategories". *Adv. Math.* 86.1, pp. 111–152. MR: 1097029 (cit. on p. 144).

Dagmar Baer (1988). "Tilting sheaves in representation theory of algebras". *Manuscripta Math.* 60.3, pp. 323–347. MR: 928291 (cit. on p. 174).

Aleksandr A. Beilinson (1978). "Coherent sheaves on $\mathbf{P}^n$ and problems in linear algebra". *Funktsional. Anal. i Prilozhen.* 12.3, pp. 68–69. MR: 509388 (cit. on p. 174).

Apostolos Beligiannis (2015). "Relative homology, higher cluster-tilting theory and categorified Auslander-Iyama correspondence". *J. Algebra* 444, pp. 367–503. MR: 3406183 (cit. on p. 167).

Petter Andreas Bergh (2014). “On the existence of cluster tilting objects in triangulated categories”. *J. Algebra* 417, pp. 1–14 (cit. on p. 158).

Raf Bocklandt (2008). “Graded Calabi Yau algebras of dimension 3”. *J. Pure Appl. Algebra* 212.1, pp. 14–32. MR: 2355031 (cit. on p. 163).

– (2012). “Consistency conditions for dimer models”. *Glasg. Math. J.* 54.2, pp. 429–447. MR: 2911380 (cit. on p. 165).

Nathan Broomhead (2012). “Dimer models and Calabi-Yau algebras”. *Mem. Amer. Math. Soc.* 215.1011, pp. viii+86. MR: 2908565 (cit. on p. 165).

Winfried Bruns and Jürgen Herzog (1993). *Cohen-Macaulay rings*. Vol. 39. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, pp. xii+403. MR: 1251956 (cit. on pp. 145, 148).

Aslak Bakke Buan, Osamu Iyama, Idun Reiten, and Jeanne Scott (2009). “Cluster structures for 2-Calabi-Yau categories and unipotent groups”. *Compos. Math.* 145.4, pp. 1035–1079. MR: 2521253 (cit. on p. 167).

Aslak Bakke Buan, Robert Marsh, Markus Reineke, Idun Reiten, and Gordana Todorov (2006). “Tilting theory and cluster combinatorics”. *Adv. Math.* 204.2, pp. 572–618. MR: 2249625 (cit. on pp. 154, 155).

Ragnar-Olaf Buchweitz (1987). “Maximal Cohen-Macaulay modules and Tate-cohomology over Gorenstein rings”. *Unpublished manuscript* (cit. on p. 148).

Ragnar-Olaf Buchweitz, Gert-Martin Greuel, and Frank-Olaf Schreyer (1987). “Cohen-Macaulay modules on hypersurface singularities. II”. *Invent. Math.* 88.1, pp. 165–182. MR: 877011 (cit. on p. 153).

Ragnar-Olaf Buchweitz and Lutz Hille (2014). “Higher Representation-Infinite Algebras from Geometry”. *Oberwolfach Rep.* 11.1, pp. 466–469 (cit. on p. 156).

Ragnar-Olaf Buchweitz, Lutz Hille, and Osamu Iyama (n.d.). *Cluster tilting for projective varieties*. In preparation (cit. on p. 175).

Ragnar-Olaf Buchweitz, Osamu Iyama, and Kota. Yamaura (Mar. 2018). *Tilting theory for Gorenstein rings in dimension one*. arXiv: 1803.05269 (cit. on pp. 144, 150, 159).

Theo Bühler (2010). “Exact categories”. *Expo. Math.* 28.1, pp. 1–69. MR: 2606234 (cit. on p. 147).

Igor Burban and Yuriy Drozd (2008). “Maximal Cohen-Macaulay modules over surface singularities”. In: *Trends in representation theory of algebras and related topics*. EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, pp. 101–166. MR: 2484725 (cit. on p. 158).

Igor Burban, Osamu Iyama, Bernhard Keller, and Idun Reiten (2008). “Cluster tilting for one-dimensional hypersurface singularities”. *Adv. Math.* 217.6, pp. 2443–2484 (cit. on p. 160).

Daniel Chan (2017). “2-hereditary algebras and almost Fano weighted surfaces”. *J. Algebra* 478, pp. 92–132. MR: 3621664 (cit. on p. 175).

Charles W. Curtis and Irving Reiner (1981). *Methods of representation theory. Vol. I*. With applications to finite groups and orders, Pure and Applied Mathematics, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, pp. xxi+819. MR: 632548 (cit. on pp. 143, 146).

Erik Darpö and Osamu Iyama (Feb. 2017). "d-representation-finite self-injective algebras". arXiv: 1702.01866 (cit. on p. 159).

Laurent Demonet and Xueyu Luo (2016). "Ice quivers with potential associated with triangulations and Cohen-Macaulay modules over orders". *Trans. Amer. Math. Soc.* 368.6, pp. 4257–4293 (cit. on p. 160).

Ernst Dieterich and Alfred Wiedemann (1986). "The Auslander-Reiten quiver of a simple curve singularity". *Trans. Amer. Math. Soc.* 294.2, pp. 455–475. MR: 825715 (cit. on p. 153).

Ju. A. Drozd and Andreĭ Vladimirovich Roĭter (1967). "Commutative rings with a finite number of indecomposable integral representations". *Izv. Akad. Nauk SSSR Ser. Mat.* 31, pp. 783–798. MR: 0220716 (cit. on p. 160).

Edgar E. Enochs and Overtoun M. G. Jenda (2000). *Relative homological algebra*. Vol. 30. De Gruyter Expositions in Mathematics. Walter de Gruyter & Co., Berlin, pp. xii+339. MR: 1753146 (cit. on pp. 145, 147).

Karin Erdmann and Thorsten Holm (2008). "Maximal n-orthogonal modules for selfinjective algebras". *Proc. Amer. Math. Soc.* 136.9, pp. 3069–3078 (cit. on p. 158).

Daniele Faenzi and Francesco Malaspina (2017). "Surfaces of minimal degree of tame representation type and mutations of Cohen-Macaulay modules". *Adv. Math.* 310, pp. 663–695. MR: 3620696 (cit. on p. 158).

Sergey Fomin and Andrei Zelevinsky (2002). "Cluster algebras. I. Foundations". *J. Amer. Math. Soc.* 15.2, pp. 497–529. MR: 1887642 (cit. on p. 154).

Robert M. Fossum, Phillip A. Griffith, and Idun Reiten (1975). *Trivial extensions of abelian categories*. Lecture Notes in Mathematics, Vol. 456. Homological algebra of trivial extensions of abelian categories with applications to ring theory. Springer-Verlag, Berlin-New York, pp. xi+122. MR: 0389981 (cit. on p. 146).

Masahiro Futaki and Kazushi Ueda (2011). "Homological mirror symmetry for Brieskorn-Pham singularities". *Selecta Math. (N.S.)* 17.2, pp. 435–452. MR: 2803848 (cit. on p. 171).

Peter Gabriel and Andreĭ Vladimirovich Roĭter (1997). *Representations of finite-dimensional algebras*. Translated from the Russian, With a chapter by B. Keller, Reprint of the 1992 English translation. Springer-Verlag, Berlin, pp. iv+177. MR: 1475926 (cit. on p. 151).

Werner Geigle and Helmut Lenzing (1987). "A class of weighted projective curves arising in representation theory of finite-dimensional algebras". In: *Singularities, representation of algebras, and vector bundles (Lambrecht, 1985)*. Vol. 1273. Lecture Notes in Math. Springer, Berlin, pp. 265–297. MR: 915180 (cit. on pp. 161, 168, 174).

– (1991). "Perpendicular categories with applications to representations and sheaves". *J. Algebra* 144.2, pp. 273–343. MR: 1140607 (cit. on pp. 161, 170).

Christof Geiss, Bernhard Keller, and Steffen Oppermann (2013). "n-angulated categories". *J. Reine Angew. Math.* 675, pp. 101–120. MR: 3021448 (cit. on p. 157).

Christof Geiss, Bernard Leclerc, and Jan Schröer (2006). "Rigid modules over preprojective algebras". *Invent. Math.* 165.3, pp. 589–632 (cit. on p. 165).

– (Mar. 2007). *Cluster algebra structures and semicanonical bases for unipotent groups*. arXiv: math/0703039 (cit. on p. 167).

– (2013). "Cluster algebras in algebraic Lie theory". *Transform. Groups* 18.1, pp. 149–178. MR: 3022762 (cit. on pp. 155, 160).

Vincent Gelinas (Feb. 2017). "Conic intersections, Maximal Cohen-Macaulay modules and the Four Subspace problem". arXiv: 1702.06608 (cit. on p. 160).

Victor Ginzburg (Dec. 2006). "Calabi-Yau algebras". arXiv: math / 0612139 (cit. on p. 162).

Shiro Goto and Kenji Nishida (2002). "Towards a theory of Bass numbers with application to Gorenstein algebras". *Colloq. Math.* 91.2, pp. 191–253. MR: 1898632 (cit. on p. 146).

Gert-Martin Greuel and Horst Knörrer (1985). "Einfache Kurvensingularitäten und torsionsfreie Moduln". *Math. Ann.* 270.3, pp. 417–425. MR: 774367 (cit. on p. 160).

Lingyan Guo (2011). "Cluster tilting objects in generalized higher cluster categories". *J. Pure Appl. Algebra* 215.9, pp. 2055–2071. MR: 2786597 (cit. on p. 155).

Dieter Happel (1988). *Triangulated categories in the representation theory of finite-dimensional algebras*. Vol. 119. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, pp. x+208. MR: 935124 (cit. on pp. 147, 148, 150–152, 158).

Martin Herschend and Osamu Iyama (n.d.). *2-hereditary algebras from hypersurfaces*. In preparation (cit. on p. 160).

– (2011). "Selfinjective quivers with potential and 2-representation-finite algebras". *Compos. Math.* 147.6, pp. 1885–1920. MR: 2862066 (cit. on p. 163).

Martin Herschend, Osamu Iyama, Hiroyuki Minamoto, and Steffen Oppermann (Sept. 2014). "Representation theory of Geigle-Lenzing complete intersections". arXiv: 1409.0668 (cit. on pp. 144, 168).

Martin Herschend, Osamu Iyama, and Steffen Oppermann (2014). "n-representation infinite algebras". *Adv. Math.* 252, pp. 292–342. MR: 3144232 (cit. on p. 156).

Daniel Huybrechts (2006). *Fourier-Mukai transforms in algebraic geometry*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, pp. viii+307. MR: 2244106 (cit. on p. 149).

Akira Ishii and Kazushi Ueda (2012). "A note on derived categories of Fermat varieties". In: *Derived categories in algebraic geometry*. EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, pp. 103–110. MR: 3050701 (cit. on p. 174).

Yasuo Iwanaga (1979). "On rings with finite self-injective dimension". *Comm. Algebra* 7.4, pp. 393–414. MR: 522552 (cit. on p. 145).

Osamu Iyama (2007a). "Auslander correspondence". *Adv. Math.* 210.1, pp. 51–82. MR: 2298820 (cit. on p. 157).

– (2007b). "Higher-dimensional Auslander-Reiten theory on maximal orthogonal subcategories". *Adv. Math.* 210.1, pp. 22–50. MR: 2298819 (cit. on pp. 155, 157).

– (2008). "Auslander-Reiten theory revisited". In: *Trends in representation theory of algebras and related topics*. EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, pp. 349–397. MR: 2484730 (cit. on p. 154).

– (2011). "Cluster tilting for higher Auslander algebras". *Adv. Math.* 226.1, pp. 1–61. MR: 2735750 (cit. on pp. 156, 165).

Osamu Iyama and Steffen Oppermann (2013). "Stable categories of higher preprojective algebras". *Adv. Math.* 244, pp. 23–68. MR: 3077865 (cit. on pp. 144, 162, 165–167).

Osamu Iyama and Ryo Takahashi (2013). "Tilting and cluster tilting for quotient singularities". *Math. Ann.* 356.3, pp. 1065–1105. MR: 3063907 (cit. on p. 165).

Osamu Iyama and Michael Wemyss (2014). "Maximal modifications and Auslander-Reiten duality for non-isolated singularities". *Invent. Math.* 197.3, pp. 521–586. MR: 3251829 (cit. on p. 146).

Osamu Iyama and Dong Yang (Feb. 2017). "Quotients of triangulated categories and Equivalences of Buchweitz, Orlov and Amiot–Guo–Keller". arXiv: 1702.04475 (cit. on p. 149).

Osamu Iyama and Yuji Yoshino (2008). "Mutation in triangulated categories and rigid Cohen-Macaulay modules". *Invent. Math.* 172.1, pp. 117–168. MR: 2385669 (cit. on p. 157).

Heinz Jacobinski (1967). "Sur les ordres commutatifs avec un nombre fini de réseaux indécomposables". *Acta Math.* 118, pp. 1–31. MR: 0212001 (cit. on p. 160).

Gustavo Jasso (2016). "n-abelian and n-exact categories". *Math. Z.* 283.3-4, pp. 703–759. MR: 3519980 (cit. on p. 157).

Bernt Tore Jensen, Alastair D. King, and Xiuping Su (2016). "A categorification of Grassmannian cluster algebras". *Proc. Lond. Math. Soc. (3)* 113.2, pp. 185–212 (cit. on p. 160).

Hiroshige Kajiura, Kyoji Saito, and Atsushi Takahashi (2007). "Matrix factorization and representations of quivers. II. Type ADE case". *Adv. Math.* 211.1, pp. 327–362. MR: 2313537 (cit. on p. 162).

– (2009). "Triangulated categories of matrix factorizations for regular systems of weights with $\epsilon = -1$". *Adv. Math.* 220.5, pp. 1602–1654. MR: 2493621 (cit. on p. 162).

Martin Kalck and Dong Yang (2016). "Relative singularity categories I: Auslander resolutions". *Adv. Math.* 301, pp. 973–1021. MR: 3539395 (cit. on p. 164).

Masaki Kashiwara and Yoshihisa Saito (1997). "Geometric construction of crystal bases". *Duke Math. J.* 89.1, pp. 9–36. MR: 1458969 (cit. on p. 160).

Bernhard Keller (1994). "Deriving DG categories". *Ann. Sci. École Norm. Sup. (4)* 27.1, pp. 63–102. MR: 1258406 (cit. on p. 154).

– (2005). "On triangulated orbit categories". *Doc. Math.* 10, pp. 551–581. MR: 2184464 (cit. on pp. 155, 164, 166).

– (2011). "Deformed Calabi-Yau completions". *J. Reine Angew. Math.* 654. With an appendix by Michel Van den Bergh, pp. 125–180. MR: 2795754 (cit. on pp. 155, 162, 163).

Bernhard Keller, Daniel Murfet, and Michel Van den Bergh (2011). "On two examples by Iyama and Yoshino". *Compos. Math.* 147.2, pp. 591–612. MR: 2776613 (cit. on p. 165).

Bernhard Keller and Idun Reiten (2008). "Acyclic Calabi-Yau categories". *Compos. Math.* 144.5. With an appendix by Michel Van den Bergh, pp. 1332–1348. MR: 2457529 (cit. on p. 165).

Bernhard Keller and Dieter Vossieck (1987). "Sous les catégories dérivées". *C. R. Acad. Sci. Paris Sér. I Math.* 305.6, pp. 225–228. MR: 907948 (cit. on p. 148).

– (1988). "Aisles in derived categories". *Bull. Soc. Math. Belg. Sér. A* 40.2. Deuxième Contact Franco-Belge en Algèbre (Faulx-les-Tombes, 1987), pp. 239–253. MR: 976638 (cit. on p. 154).

Yuta Kimura (July 2016). "Tilting and cluster tilting for preprojective algebras and Coxeter groups". arXiv: 1607.00637 (cit. on pp. 144, 154, 168).

– (2018). "Tilting theory of preprojective algebras and c-sortable elements". *J. Algebra* 503, pp. 186–221. MR: 3779993 (cit. on pp. 144, 168).

Yuta Kimura, Hiroyuki Minamoto, and Kota Yamaura (n.d.). In preparation (cit. on p. 159).

Horst Knörrer (1987). "Cohen-Macaulay modules on hypersurface singularities. I". *Invent. Math.* 88.1, pp. 153–164. MR: 877010 (cit. on p. 153).

Dirk Kussin, Helmut Lenzing, and Hagen Meltzer (2013). "Triangle singularities, ADE-chains, and weighted projective lines". *Adv. Math.* 237, pp. 194–251. MR: 3028577 (cit. on p. 171).

Helmut Lenzing and José A. de la Peña (2011). "Extended canonical algebras and Fuchsian singularities". *Math. Z.* 268.1-2, pp. 143–167. MR: 2805427 (cit. on p. 162).

Graham J. Leuschke (2012). "Non-commutative crepant resolutions: scenes from categorical geometry". In: *Progress in commutative algebra 1*. de Gruyter, Berlin, pp. 293–361. MR: 2932589 (cit. on p. 158).

Graham J. Leuschke and Roger Wiegand (2012). *Cohen-Macaulay representations*. Vol. 181. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. xviii+367. MR: 2919145 (cit. on pp. 143, 151–153).

Ming Lu and Bin Zhu (Aug. 2017). "Singularity categories of Gorenstein monomial algebras". arXiv: 1708.00311 (cit. on p. 159).

George Lusztig (1991). "Quivers, perverse sheaves, and quantized enveloping algebras". *J. Amer. Math. Soc.* 4.2, pp. 365–421. MR: 1088333 (cit. on p. 160).

Hideyuki Matsumura (1989). *Commutative ring theory*. Second. Vol. 8. Cambridge Studies in Advanced Mathematics. Translated from the Japanese by M. Reid. Cambridge University Press, Cambridge, pp. xiv+320. MR: 1011461 (cit. on p. 145).

Hiroyuki Minamoto (2012). "Ampleness of two-sided tilting complexes". *Int. Math. Res. Not. IMRN* 1, pp. 67–101. MR: 2874928 (cit. on pp. 157, 163).

Hiroyuki Minamoto and Izuru Mori (2011). "The structure of AS-Gorenstein algebras". *Adv. Math.* 226.5, pp. 4061–4095. MR: 2770441 (cit. on pp. 157, 163).

Izuru Mori and Kenta Ueyama (2016). "Stable categories of graded maximal Cohen-Macaulay modules over noncommutative quotient singularities". *Adv. Math.* 297, pp. 54–92. MR: 3498794 (cit. on p. 165).

Hiraku Nakajima (1994). "Instantons on ALE spaces, quiver varieties, and Kac-Moody algebras". *Duke Math. J.* 76.2, pp. 365–416. MR: 1302318 (cit. on p. 160).

Amnon Neeman (2001). *Triangulated categories*. Vol. 148. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. viii+449. MR: 1812507 (cit. on p. 148).

Christian Okonek, Michael Schneider, and Heinz Spindler (1980). *Vector bundles on complex projective spaces*. Vol. 3. Progress in Mathematics. Birkhauser, Boston, Mass., pp. vii+389 (cit. on p. 174).

Steffen Oppermann and Hugh Thomas (2012). "Higher-dimensional cluster combinatorics and representation theory". *J. Eur. Math. Soc. (JEMS)* 14.6, pp. 1679–1737 (cit. on p. 168).

Dmitri Orlov (2009). "Derived categories of coherent sheaves and triangulated categories of singularities". In: *Algebra, arithmetic, and geometry: in honor of Yu. I. Manin. Vol. II*. Vol. 270. Progr. Math. Birkhäuser Boston, Inc., Boston, MA, pp. 503–531. MR: 2641200 (cit. on pp. 148, 149).

Idun Reiten and Michel Van den Bergh (1989). "Two-dimensional tame and maximal orders of finite representation type". *Mem. Amer. Math. Soc.* 80.408, pp. viii+72. MR: 978602 (cit. on pp. 146, 161).

– (2002). "Noetherian hereditary abelian categories satisfying Serre duality". *J. Amer. Math. Soc.* 15.2, pp. 295–366. MR: 1887637 (cit. on pp. 149, 151).

Jeremy Rickard (1989a). "Derived categories and stable equivalence". *J. Pure Appl. Algebra* 61.3, pp. 303–317. MR: 1027750 (cit. on p. 148).

– (1989b). "Morita theory for derived categories". *J. London Math. Soc. (2)* 39.3, pp. 436–456. MR: 1002456 (cit. on p. 154).

Claus Michael Ringel (1984). *Tame algebras and integral quadratic forms*. Vol. 1099. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. xiii+376. MR: 774589 (cit. on p. 173).

Jean-Pierre Serre (1955). "Faisceaux algébriques cohérents". *Ann. of Math. (2)* 61.2, pp. 197–278 (cit. on p. 149).

Daniel Simson (1992). *Linear representations of partially ordered sets and vector space categories*. Vol. 4. Algebra, Logic and Applications. Gordon and Breach Science Publishers, Montreux, pp. xvi+499 (cit. on p. 143).

Andrzej Skowroński (2006). "Selfinjective algebras: finite and tame type". In: *Trends in representation theory of algebras and related topics*. Vol. 406. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 169–238. MR: 2258046 (cit. on p. 158).

Špela Špenko and Michel Van den Bergh (2017). "Non-commutative resolutions of quotient singularities for reductive groups". *Invent. Math.* 210.1, pp. 3–67 (cit. on p. 158).

John Tate (1957). "Homology of Noetherian rings and local rings". *Illinois J. Math.* 1, pp. 14–27. MR: 0086072 (cit. on p. 171).

Kazushi Ueda (2008). "Triangulated categories of Gorenstein cyclic quotient singularities". *Proc. Amer. Math. Soc.* 136.8, pp. 2745–2747. MR: 2399037 (cit. on p. 164).

– (2012). "On graded stable derived categories of isolated Gorenstein quotient singularities". *J. Algebra* 352, pp. 382–391. MR: 2862194 (cit. on p. 165).

Michel Van den Bergh (2004a). "Non-commutative crepant resolutions". In: *The legacy of Niels Henrik Abel*. Springer, Berlin, pp. 749–770. MR: 2077594 (cit. on pp. 153, 157).

– (2004b). "Three-dimensional flops and noncommutative rings". *Duke Math. J.* 122.3, pp. 423–455. MR: 2057015 (cit. on p. 153).

– (2015). "Calabi-Yau algebras and superpotentials". *Selecta Math. (N.S.)* 21.2, pp. 555–603. MR: 3338683 (cit. on p. 163).

Louis de Thanhoffer de Völcsey and Michel Van den Bergh (2016). "Explicit models for some stable categories of maximal Cohen-Macaulay modules". *Math. Res. Lett.* 23.5, pp. 1507–1526. MR: 3601076 (cit. on p. 164).

Michael Wemyss (2016). "Noncommutative resolutions". In: *Noncommutative algebraic geometry*. Vol. 64. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, New York, pp. 239–306. MR: 3618475 (cit. on p. 158).

Kota Yamaura (2013). "Realizing stable categories as derived categories". *Adv. Math.* 248, pp. 784–819. MR: 3107527 (cit. on pp. 144, 158).

Yuji Yoshino (1990). *Cohen-Macaulay modules over Cohen-Macaulay rings*. Vol. 146. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, pp. viii+177. MR: 1079937 (cit. on pp. 143, 150, 153).

Received 2018-01-13.

Osamu Iyama
iyama@math.nagoya-u.ac.jp
Graduate School of Mathematics
Nagoya University
Chikusa-ku, Nagoya, 464-8602
Japan

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (181–190)

ON NEGATIVE ALGEBRAIC K-GROUPS

MORITZ KERZ

Abstract

We sketch a proof of Weibel's conjecture on the vanishing of negative algebraic K-groups and we explain an analog of this result for continuous K-theory of non-archimedean algebras.

1 Negative K-groups of schemes

For a scheme X Grothendieck introduced the K-group $K_0(X)$ in his study of the generalized Riemann–Roch theorem in Berthelot, Grothendieck, and Illusie [1971, Def. IV.2.2]. In case X has an ample family of line bundles one can describe $K_0(X)$ as the free abelian group generated by the locally free $\mathcal{O}_X$-modules $\mathcal{V}$ of finite type modulo the relation $[\mathcal{V}'] + [\mathcal{V}''] - [\mathcal{V}]$ for any short exact sequence

$$0 \to \mathcal{V}' \to \mathcal{V} \to \mathcal{V}'' \to 0,$$

see Berthelot, Grothendieck, and Illusie [ibid., Sec. IV.2.9]. We denote by $X[t]$ resp. $X[t^{-1}]$ the scheme $X \times \mathbb{A}^1$ with parameter t resp. t^{-1} for the affine line $\mathbb{A}^1$, and we denote by $X[t, t^{-1}]$ the scheme $X \times \mathbb{G}_m$, where $\mathbb{G}_m = \mathbb{A}^1 \setminus \{0\}$. Bass successively defined negative algebraic K-groups of the scheme X (at least in the affine case) in degree $i < 0$ to be

$$K_i(X) = \operatorname{coker}\left[K_{i+1}(X[t]) \times K_{i+1}(X[t^{-1}]) \to K_{i+1}(X[t, t^{-1}])\right].$$

The two classical key properties, essentially due to Bass [1968], satisfied by these algebraic K-groups are the *Fundamental Theorem* and *Excision*.

The author is supported by the DFG through CRC 1085 *Higher Invariants* (Universität Regensburg).
MSC2010: primary 19D35; secondary 13D15.

Proposition 1 (Fundamental Theorem)**.** *For a quasi-compact, quasi-separated scheme X and $i \leq 0$ there exists an exact sequence*

$$0 \to K_i(X) \to K_i(X[t]) \times K_i(X[t^{-1}]) \to K_i(X[t,t^{-1}]) \to K_{i-1}(X) \to 0$$

Furthermore, for a noetherian, regular scheme X we have $K_i(X) = 0$ for $i < 0$.

Proposition 2 (Excision)**.** *For a ring homomorphism $A \to A'$ and an ideal $I \subset A$ which maps isomorphically onto an ideal I' of A' the map $K_i(A, I) \to K_i(A', I')$ of relative K-groups is an isomorphism for $i \leq 0$.*

Combining Proposition 2 with the Artin-Rees Lemma we get the following more geometric reformulation:

Corollary 3. *For a finite morphism of affine noetherian schemes $f : X' \to X$ and a closed immersion $Y \hookrightarrow X$ such that f is an isomorphism over $X \backslash Y$ the map $K_i(X, Y) \to K_i(X', Y')$ is an isomorphism for $i < 0$. Here $Y' = Y \times_X X'$.*

In general it is a hard problem to actually calculate the negative K-groups in concrete examples. One of the examples calculated in C. Weibel [2001, Sec. 6] reads:

Example 4. *For a field k and the normal surface $X = \operatorname{Spec} k[x, y, z]/(z^2 - x^3 - y^7)$ we have $K_{-1}(X) = k$ and $K_i(X) = 0$ for $i < -1$.*

In fact it is shown in C. Weibel [ibid.] that for a normal surface X we have $K_{-2}(X) = \mathbb{Z}^\rho$ and $K_i(X) = 0$ for $i < -2$, where ρ is the number of "loops" in the exceptional divisor of a resolution of singularities of X. We extend this calculation in Theorem 8 and Theorem 11 to higher dimensions.

For our results it is essential to understand in which sense we can extend Corollary 3 to global schemes. For this we have to study the non-connective algebraic K-theory spectrum $K(X)$ of a scheme X introduced in Thomason and Trobaugh [1990]. Its homotopy groups $K_i(X) = \pi_i K(X)$ for $i \leq 0$ agree with the K-groups defined above.

As shown in Thomason and Trobaugh [ibid., Sec. 8], the functor K satisfies Zariski descent. More concretely, consider a noetherian scheme X of finite dimension and a closed subscheme $Y \hookrightarrow X$. Let $K(X, Y)$ be the homotopy fibre of $K(X) \to K(Y)$. Let $K_{i,(X,Y)}$ be the Zariski presheaf on X given by $U \mapsto \pi_i K(U, Y \cap U)$ and let $K^{\sim}_{i,(X,Y)}$ be its Zariski sheafification. There exists a convergent descent spectral sequence

$$E_2^{pq} = H^p(X, K^{\sim}_{-q,(X,Y)}) \Rightarrow K_{-p-q}(X, Y). \tag{1}$$

As a direct consequence of Corollary 3 and of Zariski descent we observe:

Corollary 5. *Let X be a noetherian scheme, let $Y \hookrightarrow X$ be a closed subscheme and let d be the dimension of the closure $\overline{X \setminus Y}$. Assume that $Y \hookrightarrow X$ is an isomorphism away from $\overline{X \setminus Y}$. Let $f : \tilde{X} \to X$ be a finite morphism which is an isomorphism over $X \setminus Y$. Set $E = f^{-1}(Y)$. Then the map $f^* : K_i(X, Y) \to K_i(\tilde{X}, E)$ is an isomorphism for $i < -d$.*

Remark 6. For $\tilde{X} = X_{\text{red}}$ and $Y = \varnothing$ Corollary 5 can be refined to an isomorphism $K_i(X) \xrightarrow{\simeq} K_i(X_{\text{red}})$ for $i \leq -\dim(X)$.

Proof. In order to prove Corollary 5 one compares the descent spectral sequence (1) with the corresponding descent spectral sequence

$$E_2^{pq} = H^p(X, (f_* K_{-q,(\tilde{X},E)})^\sim) \Rightarrow K_{-p-q}(\tilde{X}, E).$$

and one uses that

(i) $K^\sim_{i,(X,Y)} \to (f_* K_{i,(\tilde{X},E)})^\sim$ is an isomorphism for $i \leq 0$ by Corollary 3,

(ii) the sheaves $K^\sim_{i,(X,Y)}$ and $(f_* K_{i,(\tilde{X},E)})^\sim$ vanish away from $\overline{X \setminus Y}$.

Note that (ii) implies that $E_2^{pq} = 0$ for $p > d$ in both spectral sequences. □

2 Platification par éclatement

In this section we explain an application of *platification par éclatement* Raynaud and Gruson [1971, Sec. 5], which generalizes the vanishing result Kerz and Strunk [2017, Prop. 5]. The motivating picture one should keep in mind is that negative K-groups of Zariski-Riemann spaces vanish, since all coherent sheaves on Zariski-Riemann spaces have Tor-dimension ≤ 1.

Let X be a quasi-compact and quasi-separated scheme, let $Y \hookrightarrow X$ be a closed subscheme defined by an invertible ideal sheaf. Recall that an admissible blow-up of X (with respect to Y) is a blow up $\mathbf{Bl}_Z X \to X$ with center $Z \hookrightarrow X$ of finite presentation and set theoretically contained in Y, see Raynaud and Gruson [1971, Def. 5.1.3]. Also recall that the composition of admissible blow-ups is admissible Raynaud and Gruson [ibid., Lem. 5.1.4]. Let $X' \to X$ be a smooth morphism of finite presentation and set $Y' = Y \times_X X'$.

The following proposition is clear in case there exists a suitable resolution of singularities for X, in view of Proposition 1. We denote by $K_i(X \text{ on } Y)$ the K-theory of X with support on Y as in Thomason and Trobaugh [1990, Def. 6.4].

Proposition 7. *Assume that X' has an ample family of line bundles and assume that X is reduced. For $i < 0$ and $\gamma \in K_i(X' \text{ on } Y')$ there exists an admissible blow-up $\tilde{X} \to X$ such that the pullback of γ to $K_i(X' \times_X \tilde{X} \text{ on } Y' \times_X \tilde{X})$ vanishes.*

Proof. For simplicity of notation we assume that $X = X'$ throughout the proof.

By noetherian approximation, see Thomason and Trobaugh [1990, App. C], there exists a directed inverse system $(X_\alpha)_\alpha$ of schemes of finite type over $\mathbb{Z}$ with affine transition maps such that $X = \varprojlim_\alpha X_\alpha$. We may further assume that Y descends to a system of closed subschemes $Y_\alpha \hookrightarrow X_\alpha$ and that there exists $\gamma_\alpha \in K_i(X_\alpha \text{ on } Y_\alpha)$ pulling back to γ.

Under the assumption that we know Proposition 7 for noetherian schemes we can, for some fixed α, find a closed subscheme Z_α which is set theoretically contained in Y_α and such that the pullback of γ_α to $K_i(\mathbf{Bl}_{Z_\alpha} X_\alpha \text{ on } Y_\alpha \times_{X_\alpha} \mathbf{Bl}_{Z_\alpha} X)$ vanishes. Let $\tilde{X}$ be $\mathbf{Bl}_Z X$, where Z is the pullback of Z_α to X. In view of the commutative diagram

$$\begin{array}{ccc} \tilde{X} & \longrightarrow & \mathbf{Bl}_{Y_\alpha} X_\alpha \\ \downarrow & & \downarrow \\ X & \longrightarrow & X_\alpha \end{array}$$

the scheme $\tilde{X}$ satisfies the requested property of Proposition 7.

By what has been explained, we can assume without loss of generality that all schemes in Proposition 7 are noetherian. In view of Bass' definition of negative K-theory, discussed in Section 1, we see that $K_{-k}(X \text{ on } Y)$ is a quotient of $K_0(\mathbb{G}^k_{m,X} \text{ on } \mathbb{G}^k_{m,Y})$ for $k > 0$ in which elements induced from $K_0(\mathbb{A}^k_X \text{ on } \mathbb{A}^k_Y)$ vanish. However, combining Kerz and Strunk [2017, Lem. 6] and Thomason and Trobaugh [1990, Ex. 5.7] we see that the latter groups are generated by coherent $\mathcal{O}$-modules on $\mathbb{G}^k_{m,X}$ resp. $\mathbb{A}^k_X$ which have support over Y and have Tor-dimension ≤ 1 over X. So without loss of generality the given element γ is induced by such an $\mathcal{O}_{\mathbb{G}^k_{m,X}}$-module $\mathcal{V}$ (here $k = -i$).

Extend $\mathcal{V}$ to a coherent $\mathcal{O}_{\mathbb{A}^k_X}$-module $\overline{\mathcal{V}}$ with support over Y. Because of the existence of an ample family of line bundles there exists an exact sequence of coherent $\mathcal{O}_{\mathbb{A}^k_X}$-modules

$$0 \to \mathcal{V}_1 \to \mathcal{V}_2 \to \overline{\mathcal{V}} \to 0$$

with $\mathcal{V}_2$ locally free. By Raynaud and Gruson [1971, Thm. 5.2.2] there exists an admissible blow-up $f : \tilde{X} \to X$ such that the strict transform of $\mathcal{V}_1$ along f is flat over X. This implies that the pullback $f^*\overline{\mathcal{V}}$ has Tor-dimension ≤ 1 over X. So the latter induces an element in $K_0(\mathbb{A}^k_{\tilde{X}})$ which induces $f^*(\gamma) \in K_i(\tilde{X})$ via the Bass construction. This shows the requested vanishing of $f^*(\gamma)$.

□

3 Weibel's conjecture

In C. A. Weibel [1980, p. 2.9] Weibel conjectured that the following Theorem 8 holds.

Theorem 8. *For a noetherian scheme X of dimension $d < \infty$ we have:*

(i) $K_i(X) = 0$ *for* $i < -d$,

(ii) $K_i(X) \xrightarrow{\simeq} K_i(X[t_1, \dots, t_r])$ *is an isomorphism for* $i \leq -d$ *and any number of variables* r.

There have been various partial results on Weibel's conjecture during the past twenty years, in particular it was shown for varieties X in characteristic zero Cortiñas, Haesemeyer, Schlichting, and C. Weibel [2008]. A complete proof of Theorem 8 was first given in Kerz, Strunk, and Tamme [2018, Thm. B] based on a pro-descent result for algebraic K-theory of blow-ups. In this section we sketch a simplified and more direct version of that proof which does not use the excision theory for K-theory of simplicial rings, as developed in Kerz, Strunk, and Tamme [ibid., Sec. 4]. For simplicity we will stick to part (i) of Theorem 8 in the proof.

Remark 9. Almost verbatim the same argument as in the proof of Theorem 8 shows that the conclusion remains true with X replaced by a scheme X' which is smooth of finite type over a noetherian scheme of dimension $d < \infty$. This was observed in Sadhu [2017].

The essential observation is that using *derived schemes* and *derived blow-ups* one can show that the analog of Corollary 5 holds for blow-ups, see Proposition 10 below.

For the convenience of the reader we summarize some properties of derived schemes in the following. A derived scheme $\mathcal{X}$ is roughly speaking given by a topological space $|\mathcal{X}|$ together with a 'derived' sheaf of commutative simplicial rings $\mathcal{O}_{\mathcal{X}}$ on $|\mathcal{X}|$, see Lurie [2016, Sec. 1.1.5]. For a derived scheme $\mathcal{X}$ its topological space together with its sheaf of homotopy groups $\pi_0 \mathcal{O}_{\mathcal{X}}$ defines an ordinary scheme, which we denote $t\mathcal{X}$. The ∞-category of derived schemes has finite limits and t preserves finite limits.

For a quasi-compact, quasi-separated derived scheme $\mathcal{X}$ one can construct its associated stable ∞-category of perfect $\mathcal{O}_{\mathcal{X}}$-modules $\mathrm{Perf}(\mathcal{X})$, see Lurie [ibid., Sec. 9.6], and one can define the K-theory spectrum $K(\mathcal{X})$ as the non-connective K-theory spectrum of $\mathrm{Perf}(\mathcal{X})$ in the sense of Blumberg, Gepner, and Tabuada [2013, Sec. 9.1].

The two key properties about the K-theory of a derived scheme $\mathcal{X}$ that we need — and that are well-known to the experts — are:

(DK1) For a quasi-compact, quasi-separated derived scheme $\mathcal{X}$ and a finite covering $\mathcal{U}$ of $\mathcal{X}$ by quasi-compact open subschemes there is a descent spectral sequence

$$E_2^{pq} = \check{H}^p(\mathcal{U}, K_{-q,\mathcal{X}}) \Rightarrow K_{-p-q}(\mathcal{X}),$$

compare Clausen, Mathew, Naumann, and Noel [2016, App. A] and Thomason and Trobaugh [1990, Prop. 8.3].

(DK2) For $\mathcal{X}$ affine, that is $\mathcal{X}$ is the spectrum of a simplicial ring, the map $K_i(\mathcal{X}) \xrightarrow{\simeq} K_i(t\mathcal{X})$ is an isomorphism for $i \leq 1$, compare Blumberg, Gepner, and Tabuada [2013, Thm. 9.53] and Kerz, Strunk, and Tamme [2018, Thm. 2.16].

Putting properties (DK1) and (DK2) together yields:

(DK3) For a quasi-compact, separated derived scheme $\mathcal{X}$ which has a covering by $d+1$ affine open subschemes the maps $K_i(\mathcal{X}) \xrightarrow{\simeq} K_i(t\mathcal{X}) \xrightarrow{\simeq} K_i((t\mathcal{X})_{\mathrm{red}})$ are isomorphisms for $i \leq -d$.

Proposition 10. *Let $X = \operatorname{Spec} A$ be a noetherian local scheme, let $Y \hookrightarrow X$ be a closed subscheme. Set $d = \dim(X)$, $\tilde{X} = \mathbf{Bl}_Y X$ and $E = f^{-1}(Y)$. Then the map $f^* : K_i(X,Y) \to K_i(\tilde{X},E)$ is an isomorphism for $i < -d$.*

Proof. Let $I \subset A$ be the ideal corresponding to Y. After replacing I by some power, we can assume that there exists a reduction of I generated by elements $a_0, \dots, a_r$ with $r < d$, see Huneke and Swanson [2006, Prop. 8.3.8]. Choose a noetherian ring A' together with a regular sequence $a'_0, \dots, a'_r \in A'$ whose image under a ring homomorphism $A' \to A$ is the sequence $a_0, \dots, a_r$. Set $X' = \operatorname{Spec} A'$ and $Y'(n) = \operatorname{Spec} A'/((a'_0)^{2^n}, \dots, (a'_r)^{2^n})$ for $n \geq 0$. The derived blow-up square

$$(2)\qquad \begin{array}{ccc} \tilde{\mathcal{X}}(m) & \longleftarrow & \mathcal{E}(m,n) \\ \downarrow & & \downarrow \\ X & \longleftarrow & \mathcal{Y}(n) \end{array}$$

is defined as the derived pullback of the usual cartesian blow-up square

$$\begin{array}{ccc} \mathbf{Bl}_{Y'(m)} X' & \longleftarrow & E'(m,n) \\ \downarrow & & \downarrow \\ X' & \longleftarrow & Y'(n) \end{array}$$

According to a derived generalization Kerz, Strunk, and Tamme [2018, Thm. 3.7] of a descent result of Thomason [1993], the square (2) gives rise to an equivalence of relative K-theory spectra $K(X, \mathcal{Y}(n)) \xrightarrow{\simeq} K(\tilde{\mathcal{X}}(n), \mathcal{E}(n,n))$ for any $n \geq 0$.

By property (DK2) above, we know that

$$K_i(X, \mathcal{Y}(n)) \xrightarrow{\simeq} K_i(X,Y)$$

is an isomorphism for $i \le 0$, $n \ge 0$ and by property (DK3) we know that

$$K_i(\tilde{\mathcal{X}}(n), \mathcal{E}(n,n)) \xrightarrow{\simeq} K_i(t\tilde{\mathcal{X}}(n), t\mathcal{E}(n,n))$$

is an isomorphism for $i < -d$, $n \ge 0$. Note that $\tilde{\mathcal{X}}$ and $\mathcal{E}$ have affine coverings by $r+1 \le d$ open subschemes.

Finally, we apply Corollary 5 to the cartesian square

$$\begin{array}{ccc} \tilde{X} & \longleftarrow & E(n) \\ \downarrow & & \downarrow \\ t\tilde{\mathcal{X}}(m) & \longleftarrow & t\mathcal{E}(m,n) \end{array}$$

for n large depending on m, in which the vertical maps are finite by Huneke and Swanson [2006, Thm. 8.2.1]. We deduce that

$$\lim_n K_i(t\tilde{\mathcal{X}}(m), t\mathcal{E}(m,n)) \xrightarrow{\simeq} \lim_n K_i(\tilde{X}, E(n))$$

is an isomorphism for $i < -d$ and $m \ge 0$. For $i < -d$ composing the isomorphisms

$$K_i(X,Y) \xrightarrow{\simeq} \lim_n K_i(X, \mathcal{Y}(n)) \xrightarrow{\simeq} \lim_n K_i(\tilde{\mathcal{X}}(n), \mathcal{E}(n,n)) \xrightarrow{\simeq} \lim_n K_i(t\tilde{\mathcal{X}}(n), t\mathcal{E}(n,n)) \xrightarrow{\simeq} \lim_m \lim_n K_i(t\tilde{\mathcal{X}}(m), t\mathcal{E}(m,n)) \xrightarrow{\simeq} \lim_n K_i(\tilde{X}, E(n)) \xrightarrow{\simeq} K_i(\tilde{X}, E)$$

finishes the proof of Proposition 10. □

Proof of Theorem 8(i). For the proof we make an induction on $d = \dim(X)$. The case $d = 0$ is clear as then $K_i(X) \xrightarrow{\simeq} K_i(X_{\mathrm{red}})$ vanishes for $i < 0$ by Proposition 1. For the induction step we use the descent spectral sequence (1) in order to reduce to the case of a local scheme $X = \operatorname{Spec} A$, see Kerz, Strunk, and Tamme [2018, Prop. 6.1] for details. Since in the affine case $K_i(X) \xrightarrow{\simeq} K_i(X_{\mathrm{red}})$ is an isomorphism for $i \le 0$, we can assume without loss of generality that X is reduced.

Fix $\gamma \in K_i(X)$ for some $i < -d$. Let $Y \hookrightarrow X$ be a closed subscheme defined by an invertible ideal sheaf such that $\gamma|_{X\setminus Y} = 0$. This means that γ can be lifted to an element $\gamma' \in K_i(X \text{ on } Y)$. By Proposition 7 there exists a blow-up $f : \tilde{X} \to X$ in a center $Z \hookrightarrow X$ which is set theoretically contained in Y such that the pullback of γ' along f vanishes, in particular $f^*(\gamma) = 0 \in K_i(\tilde{X})$.

Set $E = f^{-1}(Z)$ and consider the commutative diagram with exact rows

$$\begin{array}{ccccccc} K_{i+1}(E) & \longrightarrow & K_i(\tilde{X}, E) & \longrightarrow & K_i(\tilde{X}) & \longrightarrow & K_i(E) \\ \uparrow & & \uparrow & & \uparrow f^* & & \uparrow \\ K_{i+1}(Z) & \longrightarrow & K_i(X, Z) & \longrightarrow & K_i(X) & \longrightarrow & K_i(Z) \end{array} \tag{3}$$

As $\dim(Z), \dim(E) < d$ the K-groups in the outer corners of diagram (3) vanish by our induction assumption. The second vertical arrow in (3) is an isomorphism by Proposition 10. So the third vertical arrow is an isomorphism as well, which implies that $\gamma = 0$. □

While the negative K-groups $K_i(X)$ for $-d = -\dim(X) < i < 0$ can be quite hard to calculate, there is nice formula for $K_{-d}(X)$, which was shown in complete generality in Kerz, Strunk, and Tamme [2018, Cor. D] and previously for varieties in characteristic zero Cortiñas, Haesemeyer, Schlichting, and C. Weibel [2008].

Theorem 11. *For a noetherian scheme of dimension $d < \infty$ there is a canonical isomorphism*

$$K_{-d}(X) = H^d_{\mathrm{cdh}}(X, \mathbb{Z}),$$

where on the right we take sheaf cohomology with respect to the cdh-topology on X.

4 Negative K-groups of affinoid algebras

In this section let k be a (non-discrete) non-archimedean complete field with ring of integers k°. By π we denote an element with absolute value $0 < |\pi| < 1$. For an affinoid algebra A over k, see Bosch [2014, Sec. 3.1], we write $A\langle t\rangle$ for the Tate algebra over A, which consists of those formal power series $a_0 + a_1 t + \cdots \in A[\![t]\!]$ with $\lim_{i\to\infty} a_i = 0$, similarly for $A\langle t^{-1}\rangle$ and $A\langle t, t^{-1}\rangle$. One defines non-positive continuous K-groups of A successively by $K_0^{\mathrm{cont}}(A) = K_0(A)$ and

$$K_i^{\mathrm{cont}}(A) = \operatorname{coker}\left[K_{i+1}^{\mathrm{cont}}(A\langle t\rangle) \times K_{i+1}^{\mathrm{cont}}(A\langle t^{-1}\rangle) \to K_{i+1}^{\mathrm{cont}}(A\langle t, t^{-1}\rangle)\right]$$

for $i < 0$. These negative continuous K-groups were defined and studied in Karoubi and Villamayor [1971, Sec. 7] and Calvo [1985]. They also coincide with the continuous pro-groups defined in Morrow [2016, Sec. 3] as is shown in Kerz, Saito, and Tamme [2018, Sec. 5].

We are about to show that the analog of Weibel's conjecture, i.e. Theorem 8, holds in the non-archimedean situation:

Theorem 12. *Assume that k is discretely valued. For an affinoid k-algebra A of dimension d we have:*

(i) $K_i^{\mathrm{cont}}(A) = 0$ *for* $i < -d$,

(ii) $K_i^{\mathrm{cont}}(A) \xrightarrow{\simeq} K_i^{\mathrm{cont}}(A\langle t_1, \dots, t_r\rangle)$ *is an isomorphism for* $i \le -d$ *and any number of variables* r.

I expect that the condition that k is discretely valued in Theorem 12 can be removed. Note that even for smooth affinoid algebras A the negative continuous K-groups do not necessarily vanish, as the following example shows.

Example 13. *Assume that the residue field of* k *has characteristic zero and let* $\pi \in k$ *be an element of absolute value* $0 < |\pi| < 1$. *For the affinoid algebra* $A = k\langle s,t\rangle/(t^2 - s^3 + s^2 - \pi)$ *we have* $K_{-1}^{\mathrm{cont}}(A) = \mathbb{Z}$.

The key fact for us is that for an admissible k°-algebra A_0 in the sense of Bosch [2014, Def. 7.3.3] with $A = A_0[1/\pi]$ we obtain an exact sequence, see Kerz, Saito, and Tamme [2018, Sec. 5],

$$K_0(A_0 \text{ on } (\pi)) \to K_0((A_0/(\pi))_{\mathrm{red}}) \to K_0^{\mathrm{cont}}(A) \to K_{-1}(A_0 \text{ on } (\pi)) \to \cdots . \tag{4}$$

Recall that an admissible k°-algebra A_0 is π-adically complete, topologically of finite type and π-torsion free.

The claim made in Example 13 follows from the exact sequence (4) by setting $A_0 = k^\circ\langle s,t\rangle/(t^2 - s^3 + s^2 - \pi)$ and using that in this case $K_i(A_0 \text{ on } (\pi)) = 0$ for $i < 0$ and that $K_{-1}((A_0/(\pi))_{\mathrm{red}}) = \mathbb{Z}$.

Proof of Theorem 12(i). We can assume without loss of generality that A is reduced. An admissible blow-up of $X = \operatorname{Spec} A_0$, where A_0 is an admissible k°-algebra as above, is defined as a blow-up in a center $Y \hookrightarrow X$ which is set theoretically contained in $\operatorname{Spec} A_0/(\pi)$. Let now $\tilde{X} = \mathbf{Bl}_Y X$ be such an admissible blow-up, $X_0 = (X \otimes_{A_0} A_0/(\pi))_{\mathrm{red}}$ and $\tilde{X}_0 = (\tilde{X} \otimes_{A_0} A_0/(\pi))_{\mathrm{red}}$. For $i < -d$ we obtain from Kerz, Saito, and Tamme [ibid., Prop. 5.8] the upper exact sequence in the commutative diagram

$$\begin{array}{ccccc} K_i(\tilde{X}_0) & \longrightarrow & K_i^{\mathrm{cont}}(A) & \longrightarrow & K_{i-1}(\tilde{X} \text{ on } (\pi)) \\ \uparrow & & \| & & \uparrow \\ K_i(X_0) & \longrightarrow & K_i^{\mathrm{cont}}(A) & \longrightarrow & K_{i-1}(X \text{ on } (\pi)) \end{array} \tag{5}$$

while the lower exact sequence is just part of (4). Both groups on the left of (5) vanish by Theorem 8 since $i < -d = -\dim(X_0) = -\dim(\tilde{X}_0)$. Consider an element $\alpha \in K_i^{\mathrm{cont}}(A)$. For its image $\alpha' \in K_{i-1}(X \text{ on } (\pi))$ we can use Proposition 7 in order to choose the admissible blow-up $\tilde{X}$ such that the pullback of α' to $K_{i-1}(\tilde{X}$ on $(\pi))$ vanishes. A diagram chase in (5) shows that α vanishes. □

The forthcoming PhD thesis of C. Dahlhausen will discuss the following conjecture, which is the non-archimedean analytic variant of Theorem 11.

Conjecture 14. *For an affinoid k-algebra A of dimension d there is an isomorphism*

$$K_{-d}^{\mathrm{cont}}(A) \cong H^d(\mathfrak{M}(A), \mathbb{Z}).$$

Here $\mathfrak{M}(A)$ is the Berkovich spectrum of multiplicative seminorms Berkovich [1990, Ch. 1].

Acknowledgments. I would like to thank H. Esnault, O. Gabber, M. Groechenig, M. Morrow, S. Saito, F. Strunk, G. Tamme and C. Weibel for helpful discussions.

References

Hyman Bass (1968). *Algebraic K-theory*. W. A. Benjamin, Inc., New York-Amsterdam, pp. xx+762. MR: 0249491 (cit. on p. 181).

Vladimir G. Berkovich (1990). *Spectral theory and analytic geometry over non-Archimedean fields*. Vol. 33. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. x+169. MR: 1070709 (cit. on p. 190).

P. Berthelot, A. Grothendieck, and L. Illusie (1971). *Théorie des intersections et théorème de Riemann-Roch*. Lecture Notes in Mathematics, Vol. 225. Séminaire de Géométrie Algébrique du Bois-Marie 1966–1967 (SGA 6). Avec la collaboration de D. Ferrand, J. P. Jouanolou, O. Jussila, S. Kleiman, M. Raynaud et J. P. Serre. Springer-Verlag, Berlin-New York, pp. xii+700. MR: 0354655 (cit. on p. 181).

Andrew J. Blumberg, David Gepner, and Gonçalo Tabuada (2013). "A universal characterization of higher algebraic K-theory". *Geom. Topol.* 17.2, pp. 733–838. MR: 3070515 (cit. on pp. 185, 186).

Siegfried Bosch (2014). *Lectures on formal and rigid geometry*. Vol. 2105. Lecture Notes in Mathematics. Springer, Cham, pp. viii+254. MR: 3309387 (cit. on pp. 188, 189).

Adina Calvo (1985). "K-théorie des anneaux ultramétriques". *C. R. Acad. Sci. Paris Sér. I Math.* 300.14, pp. 459–462. MR: 788971 (cit. on p. 188).

Dustin Clausen, Akhil Mathew, Niko Naumann, and Justin Noel (June 2016). "Descent in algebraic K-theory and a conjecture of Ausoni-Rognes". arXiv: 1606.03328 (cit. on p. 186).

G. Cortiñas, C. Haesemeyer, M. Schlichting, and C. Weibel (2008). "Cyclic homology, cdh-cohomology and negative K-theory". *Ann. of Math. (2)* 167.2, pp. 549–573. MR: 2415380 (cit. on pp. 185, 188).

Craig Huneke and Irena Swanson (2006). *Integral closure of ideals, rings, and modules*. Vol. 336. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, pp. xiv+431. MR: 2266432 (cit. on pp. 186, 187).

Max Karoubi and Orlando Villamayor (1971). "K-théorie algébrique et K-théorie topologique. I". *Math. Scand.* 28, 265–307 (1972). MR: 0313360 (cit. on p. 188).

Moritz Kerz, Shuji Saito, and Georg Tamme (Feb. 2018). "K-theory of non-archimedean rings I". arXiv: 1802.09819 (cit. on pp. 188, 189).

Moritz Kerz and Florian Strunk (2017). "On the vanishing of negative homotopy K-theory". *J. Pure Appl. Algebra* 221.7, pp. 1641–1644. MR: 3614971 (cit. on pp. 183, 184).

Moritz Kerz, Florian Strunk, and Georg Tamme (2018). "Algebraic K-theory and descent for blow-ups". *Invent. Math.* 211.2, pp. 523–577. MR: 3748313 (cit. on pp. 185–188).

Jacob Lurie (2016). "Spectral algebraic geometry". Preprint (version dated October 13, 2016) (cit. on p. 185).

Matthew Morrow (Dec. 2016). "A historical overview of pro cdh descent in algebraic K-theory and its relation to rigid analytic varieties". arXiv: 1612.00418 (cit. on p. 188).

Michel Raynaud and Laurent Gruson (1971). "Critères de platitude et de projectivité. Techniques de "platification" d'un module". *Invent. Math.* 13, pp. 1–89. MR: 0308104 (cit. on pp. 183, 184).

Vivek Sadhu (Jan. 2017). "On the vanishing of Relative Negative K-theory". arXiv: 1701.09059 (cit. on p. 185).

R. W. Thomason (1993). "Les K-groupes d'un schéma éclaté et une formule d'intersection excédentaire". *Invent. Math.* 112.1, pp. 195–215. MR: 1207482 (cit. on p. 186).

R. W. Thomason and Thomas Trobaugh (1990). "Higher algebraic K-theory of schemes and of derived categories". In: *The Grothendieck Festschrift, Vol. III*. Vol. 88. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 247–435. MR: 1106918 (cit. on pp. 182–184, 186).

Charles Weibel (2001). "The negative K-theory of normal surfaces". *Duke Math. J.* 108.1, pp. 1–35. MR: 1831819 (cit. on p. 182).

Charles A. Weibel (1980). "K-theory and analytic isomorphisms". *Invent. Math.* 61.2, pp. 177–197. MR: 590161 (cit. on p. 185).

Received 2017-12-08.

Moritz Kerz
Fakultät für Mathematik
Universität Regensburg
93040 Regensburg
Germany
moritz.kerz@mathematik.uni-regensburg.de

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (191–218)

ON THE CLASSIFICATION OF FUSION CATEGORIES

Sonia Natale

Abstract

We report, from an algebraic point of view, on some methods and results on the classification problem of fusion categories over an algebraically closed field of characteristic zero.

For I could not count or name the multitude who came to Troy, though I had ten tongues and a tireless voice, and lungs of bronze as well...

Homer, *Iliad, Book II (The catalogue of ships)*

1 Introduction

Fusion categories arise from many areas of mathematics and mathematical physics encoding symmetries of structures of different nature and in this sense they can be regarded as a generalization of (finite) groups. This makes the problem of classifying fusion categories both an exciting and at the same time a colossal task. Some classes of examples of fusion categories with distinct features come from such structures like finite groups themselves, quantum groups at roots of 1, subfactors, vertex algebras... A unifying systematic approach to the theory of fusion categories was initiated in the paper Etingof, Nikshych, and Ostrik [2005]. The classification is still in an early age and perhaps awaiting for its monsters to wake up. Some progress has been made however towards the classification of fusion categories in certain classes. What we want to present here is an overview of some constructions, results and open questions related to the classification problem that we think are interesting. The world of fusion categories is a vast one and there are also important constructions, results and questions that we are not going to discuss here, mainly due to space constraints. Our approach concerns the algebraic aspect of fusion categories

The author is supported by CONICET and Secyt (UNC).
2000 *Mathematics Subject Classification:* primary 18D10; secondary 16T05.
Keywords: tensor category, fusion category, braided fusion category.

and it is for the most part motivated by different notions of extensions. The perspectives we present are the fruit of the efforts of many and the list of references at the end of the paper is not exhaustive.

We shall work over an algebraically closed base field k. Except in Sections 2.1, 2.2 and Section 4, we assume that k is of characteristic zero. We refer the reader to the book Etingof, Gelaki, Nikshych, and Ostrik [2015] and references therein for most notions on tensor and fusion categories appearing throughout.

2 Fusion categories

We start by recalling some basic definitions and notation regarding monoidal and tensor categories and the relevant functors between them. A detailed study can be found in the books Bakalov and Kirillov [2001], Etingof, Gelaki, Nikshych, and Ostrik [2015], Kassel [1995], Majid [1995], V. G. Turaev [1994].

2.1 Basic notions. A *monoidal category* is a collection $(\mathcal{C}, \otimes, \mathbf{1}, a, l, r)$, where $\mathcal{C}$ is a category, $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ is a functor, $\mathbf{1}$ is an object of $\mathcal{C}$, called the *unit object*,

$$a : \otimes \circ (\otimes \times \mathrm{id}_{\mathcal{C}}) \to \otimes \circ (\mathrm{id}_{\mathcal{C}} \times \otimes), \quad l : \mathbf{1} \otimes \mathrm{id}_{\mathcal{C}} \to \mathrm{id}_{\mathcal{C}}, \quad r : \mathrm{id}_{\mathcal{C}} \otimes \mathbf{1} \to \mathrm{id}_{\mathcal{C}},$$

are natural isomorphisms called, respectively, the associativity and left and right unit constraints, subject to the so-called *pentagon* and *triangle* axioms. For the sake of brevity, we shall simply speak of 'the monoidal category $\mathcal{C}$'.

Let $\mathcal{C}$, $\mathcal{D}$ be monoidal categories. A *monoidal functor* $\mathcal{C} \to \mathcal{D}$ is a triple (F, F^2, F^0)[1], where $F : \mathcal{C} \to \mathcal{D}$ is a functor, $F^0 : \mathbf{1} \to F(\mathbf{1})$ is an isomorphism compatible with the unit constraints, and $F^2 : \otimes \circ (F \times F) \to F \circ \otimes$ is a natural isomorphism such that, for all objects X, Y, Z of $\mathcal{C}$,

$$(F^2_{X,Y\otimes Z})(\mathrm{id}_{F(X)} \otimes F^2_{Y,Z}) a_{F(X),F(Y),F(Z)} = F(a_{X,Y,Z}) F^2_{X\otimes Y,Z}(F^2_{X,Y} \otimes \mathrm{id}_{F(Z)}).$$

An *equivalence of monoidal categories* is a monoidal functor (F, F^2, F^0) such that F is an equivalence of categories.

A monoidal category $\mathcal{C}$ is called *strict* if the associativity and unit constraints of $\mathcal{C}$ are identities. A famous result of Mac Lane states that every monoidal category is monoidally equivalent to a strict monoidal category. This allows us in (most of) what follows to suppress the associativity and unit isomorphisms.

[1] The explicit mention of F^2 and F^0 will be often omitted in what follows.

A *braiding* in a monoidal category $\mathcal{C}$ is a natural isomorphism $c_{X,Y} : X \otimes Y \to Y \otimes X$, $X, Y \in \mathcal{C}$, subject to the so-called hexagon axioms. A *braided monoidal category* is a monoidal category endowed with a braiding Joyal and Street [1993]. Braided monoidal categories such that $c_{Y,X} c_{X,Y} = \mathrm{id}_{X \otimes Y}$, for all objects $X, Y \in \mathcal{C}$, are called *symmetric*.

Let $\mathcal{C}$ be a monoidal category. Then the *Drinfeld center* $\mathcal{Z}(\mathcal{C})$ of $\mathcal{C}$ is the braided monoidal category whose objects are pairs (Z, σ_Z), where Z is an object of $\mathcal{C}$ and $\sigma_Z : Z \otimes - \to - \otimes Z$ is a natural isomorphism satisfying appropriate compatibility conditions. The tensor product of $\mathcal{Z}(\mathcal{C})$ is induced from that of $\mathcal{C}$.

Let $\mathcal{C}$ be a monoidal category. A *left dual* of an object X of $\mathcal{C}$ is a triple $(X^*, \mathrm{ev}_X, \mathrm{coev}_X)$, where X^* is an object of $\mathcal{C}$ and $\mathrm{ev}_X : X^* \otimes X \to \mathbf{1}$, $\mathrm{coev}_X : \mathbf{1} \to X \otimes X^*$, are morphisms in $\mathcal{C}$ called, respectively, evaluation and coevaluation morphisms such that the following compositions are identities:

$$X \xrightarrow{\mathrm{coev}_X \otimes \mathrm{id}_X} X \otimes X^* \otimes X \xrightarrow{\mathrm{id}_X \otimes \mathrm{ev}_X} X, \quad X^* \xrightarrow{\mathrm{id}_{X^*} \otimes \mathrm{coev}_X} X^* \otimes X \otimes X^* \xrightarrow{\mathrm{ev}_X \otimes \mathrm{id}_{X^*}} X^*,$$

A *right dual* of X is defined as a triple $({}^*X, \mathrm{ev}'_X, \mathrm{coev}'_X)$, where $\mathrm{ev}'_X : X \otimes {}^*X \to \mathbf{1}$, $\mathrm{coev}'_X : \mathbf{1} \to {}^*X \otimes X$ are morphisms satisfying similar axioms. The monoidal category $\mathcal{C}$ is called *rigid* if every object X has left and right duals X^*, ${}^*X$.

A *tensor category* over the field k is a k-linear abelian category with finite dimensional Hom spaces and objects of finite length, endowed with a rigid monoidal category structure, such that the monoidal product is k-linear in each variable and the unit object is simple. In a tensor category the monoidal product is exact in each variable. A *tensor functor* between tensor categories is a k-linear exact monoidal functor. Every tensor functor preserves duals and it is automatically faithful. A tensor functor $F : \mathcal{C} \to \mathfrak{D}$ between tensor categories $\mathcal{C}$ and $\mathfrak{D}$ is *dominant* if every object of $\mathfrak{D}$ is a subobject of $F(X)$, for some $X \in \mathcal{C}$.

A tensor category over k is called *finite* if it is equivalent as a k-linear category to the category of finite dimensional left modules over a finite dimensional k-algebra. A *fusion category over k* is a semisimple finite tensor category.

Example 2.1. Examples of tensor categories over k are given by the categories of finite dimensional left (resp. right) modules and finite dimensional left (resp. right) comodules over a *Hopf algebra* over k with bijective antipode. The tensor product in these examples is $\otimes_k$ and the associativity and unit constraints are the canonical vector space isomorphisms. These categories will be denoted, respectively, by H-mod, mod-H, H-comod, comod-H. Finite tensor categories $\mathcal{C}$ equivalent to H-mod, for some finite-dimensional Hopf algebra H, are exactly those that admit a *fiber functor*, that is, a tensor functor $\mathcal{C} \to \mathrm{Vect}_k$, where Vect_k is the tensor category of finite dimensional k-vector spaces.

More generally, if H is a *quasi-Hopf algebra* over k, then the category H-mod of finite dimensional H-modules is a tensor category over k; here the tensor product is $\otimes_k$ but the associativity constraint is induced by the associator $\Phi \in H^{\otimes 3}$ V. G. Drinfeld

[1989b]. Let H_1, H_2 be finite dimensional quasi-Hopf algebras. The tensor categories H_1-mod and H_2-mod are equivalent if and only if H_1 and H_2 are *gauge equivalent*, that is, $H_2 \cong (H_1)_F$ as quasi-Hopf algebras, where $(H_1)_F$ is certain quasi-Hopf algebra such that $(H_1)_F = H_1$ as an algebra with comultiplication $\Delta_F(h) = F\Delta(h)F^{-1}$, $h \in H_1$, and associator $\Phi_F = (1 \otimes F)(\mathrm{id} \otimes \Delta)(F)\Phi(\Delta \otimes \mathrm{id})(F^{-1})(F^{-1} \otimes 1)$.

When $\mathcal{C}$ is the representation category of a finite dimensional (quasi-)Hopf algebra H, then the Drinfeld center $\mathcal{Z}(\mathcal{C})$ is equivalent to the category $D(H)$-mod, where $D(H)$ is the quantum double of H V. G. Drinfeld [1989a], Hausser and Nill [1999], Majid [1998].

The previous example admits several generalizations. In the next example we give an outline of a construction from Hopf monads Bruguières and Virelizier [2007], Bruguières, Lack, and Virelizier [2011].

Example 2.2. A *monad* on a category $\mathcal{C}$ is an algebra in the monoidal category of endofunctors of $\mathcal{C}$. We refer the reader to Mac Lane [1998] for a study of this notion and its relation with adjunctions of functors. A *bimonad* on $\mathcal{C}$ (introduced previously by Moerdijk under the name 'Hopf monad') is a monad T endowed with a structure of a (lax) comonoidal endofunctor, that is, a natural transformation $T_2 : T \circ \otimes \to \otimes \circ (T \times T)$ and a morphism $T_0 : T(\mathbf{1}) \to \mathbf{1}$ satisfying certain compatibility conditions. A *Hopf monad* on a rigid monoidal category $\mathcal{C}$ is a bimonad equipped with a left and a right antipode Bruguières and Virelizier [2007]. If T is a Hopf monad on a rigid monoidal category $\mathcal{C}$, then the category $\mathcal{C}^T$ of T-modules in $\mathcal{C}$ is a rigid monoidal category and the forgetful functor $U : \mathcal{C}^T \to \mathcal{C}$ is a strict monoidal functor. Furthermore, suppose that $\mathcal{C}$ is a tensor category over k, and let T be a k-linear right exact Hopf monad on $\mathcal{C}$. Then $\mathcal{C}^T$ is a tensor category over k, and the forgetful functor $U : \mathcal{C}^T \to \mathcal{C}$ is a tensor functor Bruguières and Natale [2011]. Further, $\mathcal{C}^T$ is a fusion category if and only if $\mathcal{C}$ is a fusion category and T is a semisimple Hopf monad in the sense of Bruguières and Virelizier [2007].

Let $F : \mathcal{C} \to \mathfrak{D}$ be a tensor functor between tensor categories. Suppose F admits a left adjoint (which is always the case if $\mathcal{C}$ and $\mathfrak{D}$ are finite tensor categories). Since F is faithful exact, as a consequence of results of Beck on monadicity of adjunctions (see Mac Lane [1998]), there exists a Hopf monad T on $\mathfrak{D}$ such that $\mathcal{C} \cong \mathfrak{D}^T$ as tensor categories.

Let G be a finite group. The next two basic examples are special cases of those in Example 2.1.

Example 2.3. The category $\operatorname{Rep} G$ of finite dimensional k-linear representations of G is a finite tensor category over k with the usual tensor product of representations and whose unit object is the trivial representation. Thus $\operatorname{Rep} G = kG$-mod, where kG is the group (Hopf) algebra of G. By Maschke theorem, $\operatorname{Rep} G$ is a fusion category if and only if the order of G is coprime to the characteristic of k.

Two finite groups G_1 and G_2 are called *isocategorical* if the categories $\operatorname{Rep} G_1$ and $\operatorname{Rep} G_2$ are equivalent as tensor categories. This notion was introduced in Etingof and Gelaki [2001], where necessary and sufficient conditions for two finite groups to be isocategorical were given. In particular, isocategorical groups need not be isomorphic when their (common) order is divisible by 4.

Example 2.4. Let $\omega : G \times G \times G \to k$ be a 3-cocycle on G. The category Vect_G^ω of finite dimensional G-graded k-vector spaces is a fusion category with tensor product $\otimes_k$, unit object $\mathbf{1} = k$ (graded in degree $1 \in G$), and associativity constraint induced by ω. Indeed, $\mathrm{Vect}_G^\omega = H$-mod, where H is the quasi-Hopf algebra k^G of k-valued functions on G with the usual comultiplication and associator $\omega \in k^{G\times G\times G} \cong (k^G)^{\otimes 3}$. The category Vect_G^ω admits a fiber functor if and only if the class of ω is trivial in $H^3(G, k^\times)$. Equivalence classes of fusion categories of the form Vect_G^ω are in bijection with the orbit space $H^3(G, k^\times)/\operatorname{Out} G$ with respect to the natural action of the group $\operatorname{Out} G$ of outer automorphisms of G in the third cohomology group $H^3(G, k^\times)$.

Suppose that $\mathcal{C}$ is a tensor category. An object X of $\mathcal{C}$ is called *invertible* if the evaluation ev_X and the coevaluation coev_X are isomorphisms. The set G of isomorphism classes of invertible objects of $\mathcal{C}$ is a group with multiplication induced by the tensor product of $\mathcal{C}$. The tensor category $\mathcal{C}$ is called *pointed* if every simple object of $\mathcal{C}$ is invertible. Every pointed fusion category is equivalent to a category Vect_G^ω, for some 3-cocycle ω, where G is the group of invertible objects of $\mathcal{C}$.

2.2 Quantum groupoids. A *weak Hopf algebra* (or *quantum groupoid*) over k is an associative algebra H endowed with a coassociative coalgebra structure (H, Δ, ϵ) such that Δ is multiplicative, that is, $\Delta(ab) = \Delta(a)\Delta(b)$, for all $a, b \in H$, and

$$(\Delta \otimes \mathrm{id})\Delta(1) = (\Delta(1) \otimes 1)(1 \otimes \Delta(1)) = (1 \otimes \Delta(1))(\Delta(1) \otimes 1).$$
$$\epsilon(abc) = \epsilon(ab_{(1)})\epsilon(b_{(2)}c) = \epsilon(ab_{(2)})\epsilon(b_{(1)}c), \qquad \forall a, b, c \in H,$$

where $b_{(1)} \otimes b_{(2)} = \Delta(b)$. The existence of an *antipode* is also required: this is a linear map $S : H \to H$ satisfying appropriate conditions. See Böhm, Nill, and Szlachányi [1999], Bòhm and Szlachónyi [1996], Nikshych and Vainerman [2002].

A quantum groupoid H gives rise to the tensor category H-mod of its finite dimensional representations. Here, the tensor product $\otimes$ is defined as

$$V \otimes W = \Delta(1)\ V \otimes_k W,$$

for objects $V, W \in H$-mod, and by restriction of the tensor product on morphisms. The unit object is the so-called *base* subalgebra of H, and the left and right duals of an object

$V \in H$-mod are defined using the antipode. Examples of quantum groupoids from certain so-called *double groupoids* were constructed in Andruskiewitsch and Natale [2006]. Semisimple finite dimensional quantum groupoids give rise to fusion categories. The key fact about them is that every fusion category is the representation category of a quantum groupoid, in view of results of Hayashi and Ostrik:

Theorem 2.5. *Hayashi [2000], Ostrik [2003a]. Let $\mathcal{C}$ be a (multi-)fusion[2] category over k. Then there exists a finite semisimple quantum groupoid H over k such that $\mathcal{C}$ is equivalent to H-mod. Moreover, it is always possible to choose H such that its base is a commutative algebra.*

Theorem 2.5 can be generalized to finite tensor categories; in this case the result states that every finite tensor category is equivalent to H-mod for some finite dimensional left *Hopf algebroid* (a more complicated structure than a quantum groupoid that will not be discussed here). This is proved in Bruguières, Lack, and Virelizier [2011] using Hopf monads.

For the rest of this section we assume that k is the field of complex numbers.

2.3 Fusion categories from quantum groups at roots of 1. Let $\mathfrak{g}$ be a simple complex Lie algebra. Let $h^\vee$ be the dual Coxeter number of $\mathfrak{g}$ and let $q \in \mathbb{C}$ such that q^2 is a primitive root of unity of order $\ell \geq h^\vee$. We sketch here a celebrated construction, due to Andersen and Paradowski Andersen and Paradowski [1995], of a fusion category, called *Verlinde category*, associated to the pair $(\mathfrak{g}, q)$.

Let $U_q(\mathfrak{g})$ denote the Lusztig's quantized enveloping algebra specialized at q Lusztig [1993]. A $U_q(\mathfrak{g})$-module T is called a *tilting module* if both T and its dual T^* have composition series whose factors are Weyl modules. The category $\mathcal{T}$ of tilting modules, although not abelian, is a k-linear *ribbon* category (c.f. SubSection 6.1), which allows to define the trace of an endomorphism $f : T \to T$. A morphism $f : T_1 \to T_2$ in $\mathcal{T}$ is called *negligible* if $\mathrm{Tr}(fg) = 0$ for every morphism $g : T_2 \to T_1$. The *Verlinde fusion category* associated to the pair $(\mathfrak{g}, q)$ is defined as the category whose objects are tilting modules and the morphism spaces are defined by modding out negligible morphisms. We refer the reader to Andersen and Paradowski [1995], Bakalov and Kirillov [2001], Sawin [2006], V. G. Turaev [1994] and references therein for a detailed exposition about this construction.

Other celebrated construction of fusion (in fact modular) categories arising from the simple complex Lie algebra $\mathfrak{g}$ are the categories $\mathcal{C}(\mathfrak{g}, k)$ of integrable highest weight modules of level $k \in \mathbb{Z}_+$ over the corresponding affine Lie algebra $\widehat{\mathfrak{g}}$. See Bakalov and Kirillov

[2] The definition of a multi-fusion category is like that of a fusion category, but dropping the assumption of simplicity of the unit object.

[2001, Chapter 7] for a proof of this fact as well as for the relevant references and a historical overview. Alternatively, the category $\mathcal{C}(\mathfrak{g}, k)$ can be described as the category of finite length modules over the simple vertex algebra $V(\mathfrak{g}, k)$ associated with the vacuum $\widehat{\mathfrak{g}}$-module of level k Huang and Lepowsky [1999]. The relation between these categories and Verlinde categories is given by a theorem of Finkelberg [1996] that asserts that $\mathcal{C}(\mathfrak{g}, k)$ is equivalent as a modular category to the Verlinde category associated to the pair $(\mathfrak{g}, q)$, where $q = \exp(\pi i / m(k + h^\vee))$, such that $m := \langle\langle \alpha, \alpha \rangle\rangle / 2$ for a long root α of $\mathfrak{g}$, $\langle\langle\ ,\ \rangle\rangle$ being an invariant bilinear form on $\mathfrak{g}$ normalized so that $\langle\langle \beta, \beta \rangle\rangle = 2$ for short roots β. See Huang [2008].

2.4 Fusion categories from subfactors. A *subfactor* is an inclusion $A \subseteq B$ of von Neumann algebras with trivial centers. A subfactor $A \subseteq B$ is called of *finite depth* if the tensor powers of the A-bimodule B contain a finite number of isomorphism classes of simple bimodules. A subfactor $A \subseteq B$ of finite index Jones [1983] and finite depth gives rise to a (unitary) fusion category $\mathcal{C}$, called its *principal even part*: this is the full subcategory of the category of A-bimodules generated by the tensor powers of the A-bimodule B. The full subcategory of the category of B-bimodules generated by the tensor powers of the B-bimodule $B \otimes_A B$ is also a fusion category, called the *dual even part* of $A \subseteq B$, which is categorically Morita equivalent to the principal even part (c.f. SubSection 3.3).

Certain examples of fusion categories associated to subfactors do not arise from quantum groups or finite groups by means of any known contruction. Such exotic examples appear related to the Haagerup subfactor, the Asaeda-Haagerup subfactor Asaeda and Haagerup [1999], Haagerup [1994] and the extended Haagerup subfactor Bigelow, Peters, Morrison, and Snyder [2012] and have been intesively studied in the literature; see Grossman, Izumi, and Snyder [2015], Izumi [2001], Jones, Morrison, and Snyder [2014], Peters [2010], and references therein.

The Haagerup and the extended Haagerup subfactors give rise to examples of fusion categories that cannot be defined over a cyclotomic field Morrison and Snyder [2012]. Nevertheless, a result of Etingof, Nikshych, and Ostrik [2005] known as *Ocneanu rigidity* implies that every fusion category can always be defined over an algebraic number field.

3 Some invariants of a fusion category

3.1 Grothendieck ring. Let $\mathcal{C}$ be a fusion category over k and let $\mathrm{Irr}(\mathcal{C})$ denote the set of isomorphism classes of simple objects of $\mathcal{C}$. The Grothendieck group $\mathrm{Gr}(\mathcal{C})$ is the free abelian group with basis $\mathrm{Irr}(\mathcal{C})$. The cardinality of $\mathrm{Irr}(\mathcal{C})$ is called the *rank* of $\mathcal{C}$. The tensor product of $\mathcal{C}$ endows $\mathrm{Gr}(\mathcal{C})$ with a ring structure with unit element $[\mathbf{1}]$ such that, for all objects X and Y, $[X][Y] = [X \otimes Y]$, where $[X]$ denotes the isomorphism class of the object X.[3] For all $X, Y \in \operatorname{Irr}(\mathcal{C})$ we have decompositions

$$XY = \sum_{Z \in \operatorname{Irr}(\mathcal{C})} N_{X,Y}^{Z}\, Z,$$

called the *fusion rules* of $\mathcal{C}$, where the *fusion coefficients* $N_{X,Y}^{Z}$ of $\mathcal{C}$ are the non-negative integers given by $N_{X,Y}^{Z} = \dim_k \operatorname{Hom}_{\mathcal{C}}(Z, X \otimes Y)$, for all $X, Y, Z \in \operatorname{Irr}(\mathcal{C})$.

The duality of $\mathcal{C}$ induces an anti-involution of the Grothendieck ring $\operatorname{Gr}(\mathcal{C})$ that makes it into a *fusion ring* Etingof, Gelaki, Nikshych, and Ostrik [2015, Definition 3.1.3]. In particular, the fusion coefficients satisfy the relations

$$N_{X,Y}^{\mathbf{1}} = \delta_{Y,X^*}, \quad N_{X,Y}^{Z^*} = N_{Z,X}^{Y^*} = N_{Y,Z}^{X^*}, \quad \text{for all } X, Y, Z \in \operatorname{Irr}(\mathcal{C}).$$

Two fusion categories $\mathcal{C}$ and $\mathfrak{D}$ are *Grothendieck equivalent* (or *have the same fusion rules*) if there is a bijection between $\operatorname{Irr}(\mathcal{C})$ and $\operatorname{Irr}(\mathfrak{D})$ that induces a unit preserving ring isomorphism $\operatorname{Gr}(\mathcal{C}) \cong \operatorname{Gr}(\mathfrak{D})$. Non-equivalent fusion categories may share the same fusion rules: examples are the fusion categories of finite dimensional representations of the non-isomorphic non-abelian groups of order 8 Tambara and Yamagami [1998].

A *fusion subcategory* of a fusion category $\mathcal{C}$ is a full abelian subcategory closed under subquotients and tensor products. A fusion subcategory is automatically closed under duality, whence it contains the unit object of $\mathcal{C}$, and thus it is a fusion category Etingof, Gelaki, Nikshych, and Ostrik [2015, Corollary 4.11.4]. Fusion subcategories of $\mathcal{C}$ are in bijective correspondence with unital subrings of $\operatorname{Gr}(\mathcal{C})$ spanned by subsets of $\operatorname{Irr}(\mathcal{C})$. In particular, if $\mathcal{C}$ and $\mathfrak{D}$ are Grothendieck equivalent fusion categories, then the lattices of fusion subcategories of $\mathcal{C}$ and $\mathfrak{D}$ are isomorphic.

The following theorem is a consequence of a result known as *Ocneanu rigidity*, which states that fusion categories do not admit nontrivial deformations.

Theorem 3.1. *Etingof, Nikshych, and Ostrik [2005]. Up to equivalence, there is a finite number of fusion categories with a given Grothendieck ring.*

The following question remains open, although it has been established in Bruillard, Ng, E. C. Rowell, and Wang [2016b] for modular categories (see SubSection 6.1 below).

Question 3.2. Ostrik [2003b]. Are there finitely many equivalence classes of fusion categories with a given finite rank?

The answer to this question is known to be affirmative for fusion categories whose Frobenius-Perron dimension (as defined in SubSection 3.2 below) is an integer Etingof, Nikshych, and Ostrik [2005]. A related result of Etingof says that there is a finite number,

[3] By abuse of notation, we shall also write X to indicate the class of X in $\operatorname{Gr}(\mathcal{C})$.

up to Hopf algebra isomorphism, of semisimple Hopf algebras with a given finite number of irreducible representations Ostrik [2003b, Appendix]. The answer is also affirmative in small rank under certain restrictions: this was proved by Ostrik for rank 2 Ostrik [ibid.] and for rank 3 *pivotal* fusion categories Ostrik [2015].

3.2 Frobenius-Perron dimension. Let $\mathcal{C}$ be a fusion category. The *Frobenius-Perron dimension* of a simple object $X \in \mathcal{C}$ is the Frobenius-Perron eigenvalue of the matrix of left multiplication by the class of X in the basis $\operatorname{Irr}(\mathcal{C})$ of the Grothendieck ring of $\mathcal{C}$ consisting of isomorphism classes of simple objects. The *Frobenius-Perron dimension of* $\mathcal{C}$ is the number $\operatorname{FPdim}\mathcal{C} = \sum_{X\in \operatorname{Irr}(\mathcal{C})}(\operatorname{FPdim} X)^2$. The Frobenius-Perron dimension extends to a ring homomorphism $\operatorname{FPdim} : \operatorname{Gr}(\mathcal{C}) \to \mathbb{R}$ which is characterized by the fact that $\operatorname{FPdim} X > 0$, for all $X \in \operatorname{Irr}(\mathcal{C})$.

Let $\mathcal{C}$ and $\mathfrak{D}$ be fusion categories. Suppose that $F : \mathcal{C} \to \mathfrak{D}$ is a dominant tensor functor. Then the number $\operatorname{FPdim}\mathcal{C}/\operatorname{FPdim}\mathfrak{D}$ is an algebraic integer. This is also true if $\mathfrak{D}$ is a fusion subcategory of $\mathcal{C}$. See Etingof, Nikshych, and Ostrik [2005]. In particular, if $\mathcal{C}$ is integral, then $\operatorname{FPdim}\mathfrak{D}$ divides $\operatorname{FPdim}\mathcal{C}$.

Let $\mathcal{C}$ be any fusion category. For every object X of $\mathcal{C}$, $\operatorname{FPdim} X$ is a cyclotomic integer ≥ 1. Furthermore, it is known that if $\operatorname{FPdim}(X) < 2$, for some $X \in \operatorname{Irr}(\mathcal{C})$, then $\operatorname{FPdim}(X) = 2\cos(\pi/n)$, for some integer $n \geq 3$. See Etingof, Nikshych, and Ostrik [ibid.]. We have the following result on small dimensions of objects in a fusion category:

Theorem 3.3. *Calegari, Morrison, and Snyder [2011]. Let X be an object in a fusion category such that* $\operatorname{FPdim} X$ *belongs to the interval* $(2, 76/33]$. *Then* $\operatorname{FPdim} X$ *is equal to one of the following:*

$$\frac{\sqrt{7}+\sqrt{3}}{2}, \quad \sqrt{5}, \quad 1+2\cos(\frac{2\pi}{7}), \quad \frac{1+\sqrt{5}}{\sqrt{2}}, \quad \frac{1+\sqrt{13}}{2}.$$

Moreover, each of these numbers occurs as the Frobenius-Perron dimension of an object of a fusion category.

A fusion category $\mathcal{C}$ is called *integral* if $\operatorname{FPdim} X \in \mathbb{Z}$, for all simple object $X \in \mathcal{C}$. A fusion category over k is integral if and only if it is equivalent to the category of finite dimensional representations of a finite dimensional semisimple quasi-Hopf algebra over k. If this is the case, then for every H-module V we have $\operatorname{FPdim} V = \dim_k V$. Every fusion category of odd integer Frobenius-Perron dimension is integral Gelaki and Nikshych [2008].

Let $\mathcal{C}$ be an integral fusion category. One can attach some graphs to the set $\operatorname{cd}(\mathcal{C})$ of Frobenius-Perron dimensions of simple objects in the category $\mathcal{C}$: the *prime graph* $\Delta(\mathcal{C})$, whose vertices are the prime divisors of elements of $\operatorname{cd}(\mathcal{C})$ such that two vertices p and q

are joined by an edge if and only if the product pq divides some element of $\mathrm{cd}(\mathcal{C})$, and the *common divisor* graph $\Gamma(\mathcal{C})$, whose vertices are the elements of $\mathrm{cd}(\mathcal{C})-\{1\}$ such that two vertices are joined by an edge if and only if they are not coprime. These graphs extend the corresponding graphs associated to the irreducible character degrees and the conjugacy class sizes of a finite group. Some generalizations of known results on the number of connected components of these graphs for finite groups hold as well in the context of fusion categories Natale and Pacheco Rodríguez [2016].

Remark 3.4. The *categorical* or *global dimension* of a fusion category $\mathcal{C}$ is defined as $\dim \mathcal{C} = \sum_{X\in \mathrm{Irr}(\mathcal{C})} |X|^2$ where, for each simple object X of $\mathcal{C}$, $|X|^2 = \mathrm{Tr}_L(a_X)\mathrm{Tr}_L((a_X^{-1})^*)$, $a_X : X \to X^{**}$ being a fixed isomorphism (which necessarily exists in a fusion category), and $\mathrm{Tr}_L(a_X) = \mathrm{ev}_{X^*}(a_X \otimes \mathrm{id}_{X^*})\mathrm{coev}_X$ Müger [2003a]. A *spherical* structure on $\mathcal{C}$ is an isomorphism of tensor functors $\tau : \mathrm{id}_{\mathcal{C}} \to (\ )^{**}$ such that $\mathrm{Tr}_L(f) = \mathrm{Tr}_R(f)$, for any endomorphism $f : X \to X$ in $\mathcal{C}$, where Tr_L and Tr_R are certain left and right traces induced by τ. A fusion category $\mathcal{C}$ is called *pseudo-unitary* if $\dim(\mathcal{C}) = \mathrm{FPdim}(\mathcal{C})$; this is always the case if $\mathrm{FPdim}\,\mathcal{C}$ is an integer. Every pseudo-unitary fusion category $\mathcal{C}$ has a unique *spherical* structure whose categorical dimensions coincide with the Frobenius-Perron dimensions Etingof, Nikshych, and Ostrik [2005].

In the next examples we discuss some classes of non-pointed fusion categories with distinguished fusion rules.

Example 3.5. *(Near group fusion categories.)* A *near-group* fusion category is a fusion category $\mathcal{C}$ with exactly one non-invertible simple object X up to isomorphism. The fusion rules of $\mathcal{C}$ are determined by the multiplication in the group G of invertible objects of $\mathcal{C}$ and an additional relation

$$X^2 = \sum_{g\in G} g + nX, \tag{3-1}$$

where n is a non-negative integer Siehler [2003]. We say in this case that $\mathcal{C}$ has fusion rules *of type* (G, n). Near-group fusion categories of type $(1, 1)$ are called *Yang-Lee categories*: they fall into two equivalence classes and were classified by Moore and Seiberg [1989a]. Not all pairs (G, n) arise from the fusion rules of a fusion category: Ostrik proved in Ostrik [2003b] that if $n \geq 2$ then there exist no near-group fusion category with fusion rules of type $(1, n)$. Examples and results on near-group fusion categories, including restrictions on the possible values of n and its relation with the structure of G, have been obtained by Evans and Gannon and Izumi. See Evans and Gannon [2014] and references therein.

Example 3.6. *(Tambara-Yamagami fusion categories.)* A fusion category $\mathcal{C}$ is called a *Tambara-Yamagami* fusion category if $\mathcal{C}$ has near-group fusion rules of type $(G, 0)$ for some (necessarily abelian) group G. We have $\operatorname{FPdim} X = \sqrt{|G|}$ and $\operatorname{FPdim} \mathcal{C} = 2|G|$. The classification of these fusion categories was given by Tambara and Yamagami in Tambara and Yamagami [1998]: for each finite abelian group G, they are parameterized, up to equivalence, by isomorphism classes of non-degenerate symmetric bilinear forms $\chi : G \times G \to k$ and elements $\tau \in k$ such that $\tau^2 = |G|^{-1}$.

A Tambara-Yamagami fusion category is integral if and only if the order of G is a square. Tambara-Yamagami fusion categories such that G is of order 2 are called *Ising fusion categories*. These are the only non-pointed fusion categories of Frobenius-Perron dimension 4.

3.3 Categorical Morita equivalence. Let $\mathcal{C}$ be a fusion category over k. A (*right*) $\mathcal{C}$-*module category* is a finite semisimple k-linear abelian category $\mathfrak{M}$ equipped with a k-bilinear functor $\bar{\otimes} : \mathfrak{M} \times \mathcal{C} \to \mathfrak{M}$ and natural isomorphisms $\mu : \bar{\otimes} \circ (\mathrm{id}_{\mathfrak{M}} \times \otimes) \to \bar{\otimes} \circ (\bar{\otimes} \times \mathrm{id}_{\mathcal{C}})$, $r : -\bar{\otimes}\mathbf{1} \to \mathrm{id}_{\mathfrak{M}}$, satisfying certain coherence conditions similar to the pentagon and triangle axioms of a monoidal category.

Let A be an algebra in $\mathcal{C}$. Then the category ${}_A\mathcal{C}$ of left A-modules in $\mathcal{C}$ is a right $\mathcal{C}$-module category with action $\bar{\otimes} :{}_A \mathcal{C} \times \mathcal{C} \to{}_A \mathcal{C}$, given by $M \bar{\otimes} X = M \otimes X$ endowed with the natural left A-module structure. The associativity constraint of ${}_A\mathcal{C}$ is induced from that of $\mathcal{C}$.

Let $(\mathfrak{M}, \bar{\otimes})$ and $(\mathfrak{M}', \bar{\otimes}')$ be right $\mathcal{C}$-module categories. A $\mathcal{C}$-module functor $\mathfrak{M} \to \mathfrak{M}'$ is a pair (F, ζ), where $F : \mathfrak{M} \to \mathfrak{M}'$ is a functor and $\zeta_{M,X} : F(M \bar{\otimes} X) \to F(M)\bar{\otimes}' X$ is a natural isomorphism such that, for all $M \in \mathfrak{M}$, $X, Y \in \mathcal{C}$,

$$(\zeta_{M,X} \otimes \mathrm{id}_Y)\ \zeta_{M\bar{\otimes}X,Y}\ F(\mu_{M,X,Y}) = \mu'_{F(M),X,Y}\ \zeta_{M,X\otimes Y}, \quad r'_{F(M)}\ \zeta_{M,\mathbf{1}} = F(r_M).$$

An *equivalence* of $\mathcal{C}$-module categories $\mathfrak{M} \to \mathfrak{M}'$ is a $\mathcal{C}$-*module functor* $(F, \zeta) : \mathfrak{M} \to \mathfrak{M}'$ such that F is an equivalence of categories. A $\mathcal{C}$-module category is called *indecomposable* if it is not equivalent to a direct sum of two nontrivial $\mathcal{C}$-submodule categories.

Let $\mathfrak{M}, \mathfrak{M}'$ be indecomposable $\mathcal{C}$-module categories. The category $\mathrm{End}_{\mathcal{C}}(\mathfrak{M})$ of $\mathcal{C}$-module endofunctors of $\mathfrak{M}$ is a fusion category with tensor product induced by composition of functors. In particular, this gives a tool for building new examples of fusion categories from 'basic' ones. The category $\mathrm{Fun}_{\mathcal{C}}(\mathfrak{M}, \mathfrak{M}')$ of $\mathcal{C}$-module functors $\mathfrak{M} \to \mathfrak{M}'$ is an indecomposable module category over $\mathrm{End}_{\mathcal{C}}(\mathfrak{M})$ in a natural way. If A and B are indecomposable algebras in $\mathcal{C}$ such that $\mathfrak{M} \cong {}_A\mathcal{C}$ and $\mathfrak{M}' \cong {}_B\mathcal{C}$, then $\mathrm{End}_{\mathcal{C}}(\mathfrak{M})^{op}$ is equivalent to the fusion category ${}_A\mathcal{C}_A$ of (A, A)-bimodules in $\mathcal{C}$ and there is an equivalence of ${}_A\mathcal{C}_A$-module categories ${}_B\mathcal{C}_A \cong \mathrm{Fun}_{\mathcal{C}}(\mathfrak{M}, \mathfrak{M}')$, where ${}_B\mathcal{C}_A$ is the category of (B, A)-bimodules in $\mathcal{C}$.

Two fusion categories $\mathcal{C}$ and $\mathfrak{D}$ are called *categorically Morita equivalent* if $\mathfrak{D} \cong \mathrm{End}_{\mathcal{C}}(\mathfrak{M})^{op}$ for some indecomposable module category $\mathfrak{M}$. If $\mathcal{C}$ and $\mathfrak{D}$ are categorically

Morita equivalent fusion categories, then $\operatorname{FPdim} \mathcal{C} = \operatorname{FPdim} \mathcal{D}$. In addition the class of integral fusion categories is closed under categorical Morita equivalence. See Müger [2003a], Etingof, Nikshych, and Ostrik [2005]. An important characterization of the notion of categorical Morita equivalence is given by the following theorem.

Theorem 3.7. *Etingof, Nikshych, and Ostrik [2011]. Two fusion categories $\mathcal{C}$ and $\mathcal{D}$ are categorically Morita equivalent if and only if $\mathcal{Z}(\mathcal{C})$ and $\mathcal{Z}(\mathcal{D})$ are equivalent as braided fusion categories.*

4 Extensions of tensor categories

4.1 Exact sequences of tensor categories. Let $\mathcal{C}, \mathcal{D}$ be tensor categories and let $F : \mathcal{C} \to \mathcal{D}$ be a tensor functor. Let $\mathfrak{Ker}_F \subset \mathcal{C}$ denote the full subcategory of $\mathcal{C}$ whose objects are those X such that $F(X)$ is a *trivial object* of $\mathcal{D}$, that is, such that $F(X)$ is isomorphic to a direct sum of copies of the unit object $\mathbf{1}$ of $\mathcal{D}$.

A tensor functor $F : \mathcal{C} \to \mathcal{D}$ is called *normal* if for every object X of $\mathcal{C}$ there exists a subobject X_0 such that $F(X_0)$ is the largest trivial subobject of $F(X)$ in $\mathcal{D}$. The notion of normal tensor functor was introduced in Bruguières and Natale [2011] as a generalization of the notion of normal Hopf subalgebra. When $\mathcal{C}$ is a fusion category, the functor F is *normal* if and only if for every simple object $X \in \mathcal{C}$ such that $\operatorname{Hom}_{\mathcal{D}}(\mathbf{1}, F(X)) \neq 0$, we have that $X \in \mathfrak{Ker}_F$.

Definition 4.1. Bruguières and Natale [ibid.]. An *exact sequence of tensor categories* is a sequence of tensor functors

$$\mathcal{C}' \xrightarrow{i} \mathcal{C} \xrightarrow{F} \mathcal{C}'' \tag{4-1}$$

such that the functor F is dominant and normal, and i is a full embedding whose essential image is $\mathfrak{Ker}_F$. In this case we say that $\mathcal{C}$ is an extension of $\mathcal{C}''$ by $\mathcal{C}'$.

Remark 4.2. An exact sequence of tensor categories (4-1) defines a fiber functor $\mathcal{C}' \to \operatorname{Vect}_k$, since the composition $F \circ f$ maps $\mathcal{C}'$ to the trivial subcategory of $\mathcal{C}''$. By Tannaka reconstruction, (4-1) gives rise to a finite dimensional semisimple Hopf algebra H (the induced Hopf algebra of (4-1)), such that $\mathcal{C}' \cong \text{comod-}H$.

Exact sequences of tensor categories (4-1) are classified in terms of algebraic data under suitable conditions. We say that a k-linear right exact Hopf monad T on a tensor category $\mathcal{C}$ is *normal* if $T(\mathbf{1})$ is a trivial object. If T is such a Hopf monad, and if T is faithful, then it gives rise to an exact sequence of tensor categories comod-$H \to \mathcal{C}^T \to \mathcal{C}$, where, roughly, H is the Hopf algebra such that $T|_{\langle \mathbf{1} \rangle} \cong H \otimes -$ (the induced Hopf algebra of T).

Theorem 4.3. *Bruguières and Natale [ibid.]. Let $\mathcal{C}'$, $\mathcal{C}''$ be tensor categories and assume that $\mathcal{C}'$ is finite. Then the following data are equivalent:*

(i) An exact sequence (4-1);

(ii) A normal faithful k-linear right exact Hopf monad T on $\mathcal{C}''$, with induced Hopf algebra H, endowed with a tensor equivalence $K : \mathcal{C}' \cong$ comod-H.

Let $\mathcal{C}$ and $\mathfrak{D}$ be finite tensor categories. In this case exact sequences (4-1) such that F has an exact right adjoint are also classified by commutative algebras (A, σ) in the center $\mathcal{Z}(\mathcal{C})$ Bruguières and Natale [ibid., Section 6] which are *self-trivializing*, that is, such that $A \otimes A \cong A^n$ as right A-modules in $\mathcal{C}$, for some $n \geq 1$.

An exact sequence of finite tensor categories (4-1) is called *central* if, denoting by (A, σ) the corresponding commutative algebra in $\mathcal{Z}(\mathcal{C})$, the tensor functor $i : \mathcal{C}' \to \mathcal{C}$ lifts to a tensor functor $\tilde{i} : \mathcal{C}' \to \mathcal{Z}(\mathcal{C})$ such that $\tilde{i}(A) = (A, \sigma)$ Bruguières and Natale [2014].

4.2 Hopf algebra extensions. Let H, H', H'' be Hopf algebras over the field k. A sequence of Hopf algebra maps

$$k \longrightarrow H' \xrightarrow{i} H \xrightarrow{f} H'' \longrightarrow k, \tag{4-2}$$

is called a (strictly) *exact sequence of Hopf algebras* if i is injective, π is surjective, $i(H') = H^{co\pi} = \{h \in H : (\mathrm{id} \otimes \pi)\Delta(h) = h \otimes 1\}$, and H is right faithfully flat over $i(H')$ (the last condition is automatic in the finite dimensional case). Letting $\iota = i_*$ and $F = f_*$ to be the functors induced by restriction along i and f, respectively, the exact sequence (4-2) induces an exact sequence of tensor categories

$$\text{comod-}H' \xrightarrow{\iota} \text{comod-}H \xrightarrow{F} \text{comod-}H''.$$

Suppose H is finite dimensional. If (4-2) is an exact sequence, then $i(H')$ is a *normal* Hopf subalgebra of H, that is, a Hopf subalgebra stable under the adjoint actions of H. Conversely, every normal Hopf subalgebra H' of H gives rise to an exact sequence (4-2), where $i : H' \longrightarrow H$ is the inclusion and $f : H \longrightarrow H/H(H')^+ =: H''$ is the canonical projection. In this case H can be recovered as a bicrossed product $H' \# H''$ with respect to suitable cohomological data Andruskiewitsch and Devoto [1995]. A Hopf algebra is called *simple* if it contains no proper normal Hopf subalgebra. The following result implies that the simplicity of a Hopf algebra is not a categorical notion.

Theorem 4.4. *Galindo and Natale [2007]. There exists a simple Hopf algebra H such that H-mod $\cong$* Rep G*, where G is a solvable group.*

In fact in the examples of Galindo and Natale [2007], one may take H of dimension p^2q^2, where p and q are distinct prime numbers, thus showing that the analogue of Burnside's p^aq^b-Theorem does not extend to the context of semisimple Hopf algebras with the natural definition of normal Hopf subalgebra. In the (characteristic zero) semisimple case, the only simple examples with dimension ≤ 60 arise in dimensions 36 and 60 and they are twistings of group algebras Natale [2007], Natale [2010].

Composition series of a finite dimensional Hopf algebra H were introduced by Andruskiewitsch and Müller as a sequence of finite dimensional simple Hopf algebras $\mathfrak{H}_1, \dots, \mathfrak{H}_n$ defined recursively as follows: If H is simple, then $n = 1$ and $\mathfrak{H}_1 = H$. If, on the other hand, $k \subsetneq A \subsetneq H$ is a normal Hopf subalgebra, and $\mathfrak{A}_1, \dots, \mathfrak{A}_m$, $\mathfrak{B}_1, \dots, \mathfrak{B}_l$, are composition series of A and $B = H/HA^+$, respectively, then we let $n = m + l$ and $\mathfrak{H}_i = \mathfrak{A}_i$, if $1 \leq i \leq m$, $\mathfrak{H}_i = \mathfrak{B}_{i-m}$, if $m < i \leq m + l$.

The following analogue of the Jordan-Hölder theorem holds for finite dimensional Hopf algebras; see Andruskiewitsch [2002, Question 2.1].

Theorem 4.5. *Natale [2015]. Let $\mathfrak{H}_1, \dots, \mathfrak{H}_n$ and $\mathfrak{H}'_1, \dots, \mathfrak{H}'_m$ be two composition series of H. Then there exists a bijection $f : \{1, \dots, n\} \to \{1, \dots, m\}$ such that $\mathfrak{H}_i \cong \mathfrak{H}'_{f(i)}$ as Hopf algebras.*

4.3 Exact sequences with respect to a module category. The notion of exact sequence of tensor categories was generalized in Etingof and Gelaki [2017] to that of exact sequence of finite tensor categories with respect to a module category.

Let $\mathfrak{C}$ and $\mathfrak{D}$ be finite k-linear abelian categories. Their *Deligne tensor product* is a finite tensor category denoted $\mathfrak{C} \boxtimes \mathfrak{D}$ endowed with a functor $\boxtimes : \mathfrak{C} \times \mathfrak{D} \to \mathfrak{C} \boxtimes \mathfrak{D}$ exact in both variables such that for any k-bilinear right exact functor $F : \mathfrak{C} \times \mathfrak{D} \to \mathfrak{A}$, where $\mathfrak{A}$ is a k-linear abelian category, there exists a unique right exact functor $\tilde{F} : \mathfrak{C} \boxtimes \mathfrak{D} \to \mathfrak{A}$ such that $\tilde{F} \circ \boxtimes = F$. Such a category exists and it is unique up to equivalences. In fact, if $\mathfrak{C} \cong A$-mod and $\mathfrak{D} \cong B$-mod, for some finite dimensional k-algebras A and B, then $\mathfrak{C} \boxtimes \mathfrak{D} \cong (A \otimes B)$-mod. See Deligne [1990]. The tensor product of two finite tensor categories $\mathfrak{C}$ and $\mathfrak{D}$ is again a finite tensor category and if $\mathfrak{C}$ and $\mathfrak{D}$ are fusion categories, then so is $\mathfrak{C} \boxtimes \mathfrak{D}$.

Let $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{C}$ be finite tensor categories and let $\mathfrak{M}$ be an exact indecomposable left $\mathfrak{A}$-module category[4]; in particular, $\mathfrak{M}$ is finite. Let $\operatorname{End}(\mathfrak{M})$ denote the category of k-linear right exact endofunctors of $\mathfrak{M}$, which is a monoidal category with tensor product given by composition of functors. Let also $i : \mathfrak{A} \to \mathfrak{B}$ denote the inclusion functor.

[4]Exactness of $\mathfrak{M}$ means that $P \otimes M$ is projective for any projective object $P \in \mathfrak{A}$ and any object $M \in \mathfrak{M}$.

Definition 4.6. Etingof and Gelaki [2017]. An exact sequence of tensor categories *with respect to* $\mathfrak{M}$ is a sequence of tensor functors

$$\mathfrak{A} \xrightarrow{i} \mathfrak{B} \xrightarrow{F} \mathfrak{C} \boxtimes \operatorname{End}(\mathfrak{M}),$$

such that F is dominant, $\mathfrak{A}$ coincides with the subcategory of $\mathfrak{B}$ mapped to $\operatorname{End}(\mathfrak{M})$ under F and, for every object X of $\mathfrak{B}$, there exists a subobject X_0 of X such that $F(X_0)$ is the largest subobject of $F(X)$ contained in $\operatorname{End}(\mathfrak{M})$. In this case $\mathfrak{B}$ is said an extension of $\mathfrak{C}$ by $\mathfrak{A}$ with respect to $\mathfrak{M}$.

It was shown in Etingof and Gelaki [ibid.] that the Deligne tensor product of two tensor categories gives rise to an exact sequence in the sense of the previous definition. The notion of exact sequence with respect to a module category is self-dual in an appropriate sense. In addition, if $\mathfrak{A}$ and $\mathfrak{C}$ are fusion categories and $\mathfrak{M}$ is an indecomposable exact (thus semisimple) module category over $\mathfrak{A}$, then any extension of $\mathfrak{C}$ by $\mathfrak{A}$ with respect to $\mathfrak{M}$ is also a fusion category.

The answers to the following natural questions are at the moment not known:

Question 4.7. Does an analogue of the Jordan-Hölder theorem hold for finite tensor (fusion) categories? Is it possible to classify *simple* tensor (fusion) categories?

5 Fusion categories from finite groups

A fusion category $\mathfrak{C}$ is called *group-theoretical* if it is categorically Morita equivalent to a pointed fusion category. Let $\mathfrak{C}$ be a pointed fusion category, so that there exist a finite group G and a 3-cocycle $\omega : G \times G \times G \to k^\times$ such that $\mathfrak{C} \cong \operatorname{Vect}_G^\omega$, c.f. Example 2.4. Every indecomposable module category over $\operatorname{Vect}_G^\omega$ arises from a pair (F, α), where F is a subgroup of G and $\alpha : F \times F \to k^\times$ is a 2-cochain on F such that $d\alpha = \omega|_{F \times F \times F}$. Thus, the restriction $\omega|_F$ represents the trivial cohomology class in $H^3(F, k^\times)$. The (left) module category associated to such pair (F, α) is the category $\mathfrak{M}_0(F, \alpha) = (\operatorname{Vect}_G^\omega)_{k_\alpha F}$ of (right) $k_\alpha F$-modules in $\operatorname{Vect}_G^\omega$, where $k_\alpha F$ is the twisted group algebra of F. The group-theoretical category $(\operatorname{Vect}_G^\omega)^*_{\mathfrak{M}_0(F,\alpha)}$ is denoted $\mathfrak{C}(G, \omega, F, \alpha)$. Every group-theoretical fusion category is integral. Necessary and sufficient conditions for a group-theoretical category to be equivalent to the representation category of a (semisimple) Hopf algebra were given in Ostrik [2003c].

The class of group-theoretical fusion categories is quite well-understood. It is almost tautologically closed under categorical Morita equivalence, as well as under Deligne tensor products and Drinfeld centers. Moreover, a fusion category $\mathfrak{C}$ is equivalent to a group-theoretical fusion category $\mathfrak{C}(G, \omega, F, \alpha)$ if and only if its Drinfeld center $\mathcal{Z}(\mathfrak{C})$ is equivalent as a braided fusion category to the category of finite-dimensional representations of the *twisted quantum double* $D^\omega G$ Dijkgraaf, Pasquier, and Roche [1991], Majid [1998].

The next theorem illustrates that certain restrictions on the dimension of an integral fusion category force it to be group-theoretical.

Theorem 5.1. *Let p, q, r be distinct prime numbers. Then:*

(i) *V. Drinfeld, Gelaki, Nikshych, and Ostrik [2007]. Every integral fusion category of Frobenius-Perron dimension p^n, $n \geq 0$, is group-theoretical.*

(ii) *Etingof, Gelaki, and Ostrik [2004], Etingof, Nikshych, and Ostrik [2011]. Every integral fusion category of Frobenius-Perron dimension pq or pqr is group-theoretical.*

Example 5.2. *(Abelian extensions of Hopf algebras.)* Exact sequences of finite dimensional Hopf algebras (4-2) such that $H' \cong k^\Gamma$ and $H'' \cong kF$, for some finite groups F and Γ, are called *abelian extensions*. An abelian extension of kF by k^Γ arises from mutual actions by permutations $\Gamma \overset{\triangleleft}{\leftarrow} \Gamma \times F \overset{\triangleright}{\rightarrow} F$ that make (F, Γ) into a *matched pair* of finite groups. This amounts to the existence of a group G together with an exact factorization $G = F\Gamma$: the actions $\triangleleft, \triangleright$ are this case determined by the relations $gx = (g \triangleright x)(g \triangleleft x)$, for all $x \in F$, $g \in \Gamma$.

Let (F, Γ) be a matched pair of finite groups. Let also $\sigma : F \times F \to (k^*)^\Gamma$ and $\tau : \Gamma \times \Gamma \to (k^*)^F$ be normalized 2-cocycles. Under suitable conditions, one can associate a Hopf algebra $H = k^\Gamma {}^\tau\#_\sigma kF$ (with crossed product algebra structure and crossed coproduct coalgebra structure) that fits into an exact sequence of Hopf algebras $k \to k^\Gamma \to H \to kF \to k$. Moreover, every Hopf algebra H fitting into an exact sequence of this form is isomorphic to $k^\Gamma {}^\tau\#_\sigma kF$ for appropriate data $\triangleleft, \triangleright, \sigma, \tau$. Equivalence classes of such extensions associated to a fixed pair $(\triangleleft, \triangleright)$ form an abelian group whose unit element is the class of the *split* extension $k^\Gamma \# kF$.

Abelian extensions are among the first non-commutative and non-cocommutative examples of Hopf algebras in the literature; they were studied by G.I. Kac in the 60's Kac [1962]. We refer to Masuoka [2002] for results on the cohomology underlying an abelian exact sequence and generalizations.

Every abelian extension is group-theoretical. Indeed, if $H \cong k^\Gamma {}^\tau\#_\sigma kF$, then there is equivalence of tensor categories mod-$H \cong \mathcal{C}(G, \omega, F, 1)$, where $\omega : G \times G \times G \to k^\times$ is a 3-cocycle coming from the class of H in an exact sequence due to G. I. Kac. However, group-theoretical Hopf algebras are not closed under taking extensions. Examples of semisimple Hopf algebras H which are not group-theoretical were constructed by Nikshych, answering a question of Etingof, Nikshych, and Ostrik [2005]:

Theorem 5.3. *Nikshych [2008]. There exist semisimple Hopf algebras which are not group-theoretical.*

The examples of Nikshych [ibid.] have dimension $4p^2$, where p is an odd prime number. These Hopf algebras H fit into a central exact sequence $k \to k^{\mathbb{Z}_2} \to H \to A \to k$, where A is certain group-theoretical Hopf algebra of dimension $2p^2$. The 36 dimensional example arising from Nikshych [ibid.] gives the smallest semisimple Hopf algebra which is not group-theoretical.

5.1 Group extensions and equivariantization. Let G be a finite group. A G-grading on a fusion category $\mathcal{C}$ is a decomposition $\mathcal{C} = \oplus_{g \in G} \mathcal{C}_g$, such that $\mathcal{C}_g \otimes \mathcal{C}_h \subseteq \mathcal{C}_{gh}$, for all $g, h \in G$. The fusion category $\mathcal{C}$ is called a *G-extension* of a fusion category $\mathfrak{D}$ if there is a faithful grading $\mathcal{C} = \oplus_{g \in G} \mathcal{C}_g$ with neutral component $\mathcal{C}_1 \cong \mathfrak{D}$. Group extensions of a fusion category were classified in Etingof, Nikshych, and Ostrik [2010] in terms of various data related to some low degree cohomology groups.

If $\mathcal{C}$ is any fusion category, there exist a finite group $U(\mathcal{C})$, called the *universal grading group* of $\mathcal{C}$, and a canonical faithful grading $\mathcal{C} = \oplus_{g \in U(\mathcal{C})} \mathcal{C}_g$, with neutral component $\mathcal{C}_{ad}$, where $\mathcal{C}_{ad}$ is the adjoint subcategory of $\mathcal{C}$, that is, the fusion subcategory generated by $X \otimes X^*$, $X \in \operatorname{Irr}(\mathcal{C})$. In addition, if $\operatorname{FPdim} \mathcal{C} \in \mathbb{Z}$, then $\mathcal{C}$ is faithfully graded by an elementary abelian 2-group E. Moreover, there is a set of distinct square-free integers n_g, $g \in E$, such that $n_1 = 1$ and $\operatorname{FPdim} X \in \mathbb{Z}\sqrt{n_g}$, for every simple object X of $\mathcal{C}_g$. The neutral component of this grading is the unique maximal integral fusion subcategory of $\mathcal{C}$. See Gelaki and Nikshych [2008].

A fusion category $\mathcal{C}$ is (cyclically) *nilpotent* if there exists a sequence of fusion categories $\operatorname{Vect}_k = \mathcal{C}_0 \subseteq \mathcal{C}_1 \cdots \subseteq \mathcal{C}_n = \mathcal{C}$, and finite (cyclic) groups $G_1, \dots, G_n$, such that for all $i = 1, \dots, n$, $\mathcal{C}_i$ is a G_i-extension of $\mathcal{C}_{i-1}$.

Another kind of 'group extension' of a fusion category is provided by the equivariantization under a finite group action. An *action* of a finite group G on a fusion category $\mathcal{C}$ by tensor autoequivalences is a monoidal functor $\rho : \underline{G} \to \operatorname{Aut}_\otimes \mathcal{C}$, where $\underline{G}$ is the strict monoidal category with objects $g \in G$, identities as its only morphisms, and group multiplication as tensor product, and $\operatorname{Aut}_\otimes \mathcal{C}$ is the monoidal category of tensor autoequivalences of $\mathcal{C}$ where morphisms are isomorphisms of tensor functors. The *equivariantization* of $\mathcal{C}$ with respect to the action ρ, denoted $\mathcal{C}^G$, is a fusion category whose objects are pairs (X, μ), such that X is an object of $\mathcal{C}$ and $\mu = (\mu^g)_{g \in G}$, is a collection of isomorphisms $\mu^g : \rho^g X \to X$, $g \in G$, satisfying

$$\mu^g \rho^g(\mu^h) = \mu^{gh} \rho^{g,h}_{2_X}, \quad \mu_1 \rho_{0_X} = \mathrm{id}_X,$$

for all $g, h \in G$, where ρ^2 and ρ^0 denote the monoidal structure of ρ.

Remark 5.4. If $\rho : \underline{G} \to \operatorname{Aut}_\otimes \mathcal{C}$ is an action of a group G by tensor autoequivalences, then the forgetful functor $F : \mathcal{C}^G \to \mathcal{C}$, $F(X, \mu) = X$, is a normal dominant tensor functor and it gives rise to a central exact sequence of fusion categories $\operatorname{Rep} G \to \mathcal{C}^G \to \mathcal{C}$

Bruguières and Natale [2014]. Thus equivariantization provides examples of exact sequences of tensor categories. The notion of equivariantization explained above can be extended to define equivariantization of tensor categories under finite group scheme actions. Some characterizations of an exact sequence of tensor categories arising from equivariantization where given in Bruguières and Natale [ibid.]. In particular, an exact sequence of finite tensor categories arises from an equivariantization if and only if the sequence is central (see SubSection 4.1), which is an alternative formulation of a previous characterization given in Etingof, Nikshych, and Ostrik [2011] in the context of fusion categories.

A generalization of the equivariantization construction was given in Natale [2016b]: the input for this construction is a *crossed action* of a matched pair of finite groups (G, Γ) in a tensor category $\mathcal{C}$, which means a k-linear action of G on $\mathcal{C}$ (not necessarily by tensor autoequivalences) and a Γ-grading on $\mathcal{C}$ satisfying certain compatibility conditions. The resulting tensor category $\mathcal{C}^{(G,\Gamma)}$ also fits into an exact sequence $\operatorname{Rep} G \to \mathcal{C}^{(G,\Gamma)} \to \mathcal{C}$. The representation categories of abelian extensions (c.f. Example 5.2) are also contained in this construction.

Associated to an action of a group on a fusion category, there is another fusion category, called a *crossed product*, and denoted $\mathcal{C} \rtimes G$ Tambara [2001]. As a k-linear abelian category $\mathcal{C} \rtimes G = \mathcal{C} \boxtimes \operatorname{Vect}_G$ with tensor product defined by

$$(X \boxtimes g) \otimes (Y \boxtimes h) = (X \otimes \rho^g(Y)) \boxtimes gh,$$

for all $X, Y \in \mathcal{C}$, $g, h \in G$, unit object $\mathbf{1} \boxtimes k$ and associativity and unit constraints induced from those of $\mathcal{C}$. The category $\mathcal{C} \rtimes G$ is a G-extension of $\mathcal{C}$ with homogeneous components $(\mathcal{C} \rtimes G)_g = \mathcal{C} \boxtimes g$, $g \in G$. The relation with the equivariantized fusion category is given by an equivalence of tensor categories $(\mathcal{C}^G)^*_{\mathcal{C}} \cong \mathcal{C} \rtimes G$ Nikshych [2008]. In particular, every equivariantization is categorically Morita equivalent to a (graded) group extension.

Remark 5.5. The representation category of the non-group-theoretical examples of Nikshych are constructed in Nikshych [ibid.] as an equivariantization of a Tambara-Yamagami category of dimension $2p^2$ under the action of the group $\mathbb{Z}_2$. Thus, *a fortiori*, the class of group-theoretical fusion categories is not closed under group equivariantizations and neither under group extensions. Necessary and sufficient conditions for a group extension of a fusion category to be group-theoretical were given in Nikshych [ibid.].

Example 5.6. (*Equivariantization of pointed fusion categories.*) Let $\mathcal{C} = \operatorname{Vect}_\Gamma^\omega$ be a pointed fusion category, where Γ is a finite group and $\omega : \Gamma\times\Gamma\times\Gamma \to k^\times$ is a normalized 3-cocycle. Let also G be a finite group. An action $\rho : \underline{G} \to \operatorname{Aut}_\otimes \mathcal{C}$ corresponds to an action by group automorphisms of G on Γ, $x \mapsto {}^g x$, $x \in \Gamma$, $g \in G$ and maps $\tau : G\times\Gamma\times\Gamma \to k^\times$,

$\sigma : G \times G \times \Gamma \to k^\times$, obeying, for all $x, y, z \in \Gamma$, $g, h, l \in G$, the following conditions:

$$\sigma(h,l;x)\,\sigma(g,hl;x) = \sigma(gh,l;x)\,\sigma(g,h;{}^{l}x)$$
$$\frac{\omega(x,y,z)}{\omega({}^{g}x,{}^{g}y,{}^{g}z)} = \frac{\tau(g;xy,z)\tau(g;x,y)}{\tau(g;y,z)\,\tau(g;x,yz)}$$
$$\frac{\tau(gh;x,y)}{\tau(g;{}^{h}x,{}^{h}y)\,\tau(h;x,y)} = \frac{\sigma(g,h;x)\,\sigma(g,h;y)}{\sigma(g,h;xy)}$$

In this example, the category $\mathcal{C} \rtimes G$ is pointed: indeed, $\mathcal{C} \rtimes G \cong \mathrm{Vect}^{\tilde{\omega}}_{\Gamma \rtimes G}$, where $\Gamma \rtimes G$ is the semidirect product associated to the given action by group automorphisms of G on Γ and $\tilde{\omega}$ is a certain 3-cocycle on $\Gamma \rtimes G$ Tambara [2001]. In particular, any equivariantization of a pointed fusion category is group-theoretical.

5.2 Weakly group-theoretical fusion categories. A fusion category $\mathcal{C}$ is called *weakly group-theoretical* (respectively, *solvable*) if it is categorically Morita equivalent to a nilpotent (respectively, cyclically nilpotent) fusion category Etingof, Nikshych, and Ostrik [2011]. Weakly group-theoretical fusion categories can be described by means of group-theoretical data. The notion of solvability extends that of finite groups: the fusion categories $\mathrm{Rep}\, G$ and $\mathrm{Vect}^{\omega}_{G}$ are solvable if and only if G is solvable. However, not every nilpotent fusion category is solvable: for instance, Vect_G is always nilpotent.

The class of weakly group-theoretical fusion categories is stable under the operations of taking extensions, equivariantizations, Morita equivalent categories, tensor products, Drinfeld center, fusion subcategories and components of quotient categories. Also, the class of solvable fusion categories is stable under taking extensions and equivariantizations by solvable groups, Morita equivalent categories, tensor products, Drinfeld center, fusion subcategories and components of quotient categories. See Etingof, Nikshych, and Ostrik [ibid.].

Every weakly group-theoretical fusion category has integer Frobenius-Perron dimension. The following is the most important open question regarding the classification of this class of fusion categories:

Question 5.7. Etingof, Nikshych, and Ostrik [ibid.]. Is every fusion category with integer Frobenius-Perron dimension weakly group-theoretical?

A related open question is the following:

Question 5.8. Is the class of weakly group-theoretical fusion categories closed under extensions?

We summarize in the next theorem some results related to Question 5.7. Part (ii) is a generalization of a well-known theorem of Bunrside for finite groups.

Theorem 5.9. *Let $\mathcal{C}$ be a fusion category. Then the following hold:*

(i) *Etingof, Nikshych, and Ostrik [2005]. If* $\operatorname{FPdim} \mathcal{C} = p^n$, *p prime,* $n \geq 0$, *then* $\mathcal{C}$ *is nilpotent.*

(ii) *Etingof, Nikshych, and Ostrik [2011]. If* $\operatorname{FPdim} \mathcal{C} = p^n q^m$, *p, q primes,* $n, m \geq 0$, *then* $\mathcal{C}$ *is solvable.*

(iii) *Natale [2014]. If* $\mathbb{C}$ *is braided non-degenerate and* $\operatorname{FPdim} \mathcal{C} = dq^n$, *p prime,* $n \geq 0$ *and d a square-free integer, then* $\mathcal{C}$ *is solvable.*

It was shown in Etingof, Nikshych, and Ostrik [2011] that a fusion category $\mathcal{C}$ is weakly group-theoretical if and only if there exists a series of fusion categories

$$\text{Vect} = \mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_n = \mathcal{C}, \tag{5-1}$$

such that for all $1 \leq i \leq n$, the Drinfeld center $\mathcal{Z}(\mathcal{C}_i)$ contains a Tannakian subcategory $\mathcal{E}_i$ and the de-equivariantization of the Müger centralizer $\mathcal{E}_i'$ in $\mathcal{Z}(\mathcal{C}_i)$ by $\mathcal{E}_i$ is equivalent to $\mathcal{Z}(\mathcal{C}_{i-1})$ as braided fusion categories. Since the categories $\mathcal{E}_i$ are Tannakian, then, for all $i = 1, \dots, n$, there exist finite groups $G_1, \dots, G_n$, such that $\mathcal{E}_i \cong \operatorname{Rep} G_i$ as symmetric fusion categories (c.f. Section 6 below). A *composition series* of $\mathcal{C}$ is a series (5-1) whose *factors* $G_1, \dots, G_n$, are simple groups.

The following theorem is an analogue of the Jordan-Hölder theorem. Its proof relies on the structure of a crossed braided fusion category (c.f. SubSection 6.3).

Theorem 5.10. *Natale [2016a]. Let* $\mathcal{C}$ *be a weakly group-theoretical fusion category. Then two composition series of* $\mathcal{C}$ *have, up to isomorphisms, the same factors counted with multiplicities. Thus they are invariants of* $\mathcal{C}$ *under categorical Morita equivalence.*

The solvability of a finite group G is known to be determined by its character table or, equivalently, by the fusion rules of the category $\operatorname{Rep} G$. The answer to the analogous question for fusion categories is not known.

Question 5.11. Escañuela González and Natale [2017]. Is the solvability of a fusion category determined by its fusion rules?

6 Braided fusion categories

Let $\mathcal{C}$ be a braided fusion category and let $\mathfrak{D}$ be a fusion subcategory of $\mathcal{C}$. The *Müger centralizer* of $\mathfrak{D}$ in $\mathcal{C}$, denoted $\mathfrak{D}'$, is the full fusion subcategory generated by all objects $X \in \mathcal{C}$ such that $c_{Y,X} c_{X,Y} = \mathrm{id}_{X \otimes Y}$, for all objects $Y \in \mathfrak{D}$. The category $\mathcal{C}'$ is called the *Müger center* (or *symmetric center*) of $\mathcal{C}$. If $\mathcal{C}$ is any braided fusion category, its Müger center $\mathcal{C}'$ is a symmetric fusion subcategory of $\mathcal{C}$. On the opposite extreme, $\mathcal{C}$ is called *non-degenerate* if $\mathcal{C}' \cong \text{Vect}$. See V. Drinfeld, Gelaki, Nikshych, and Ostrik [2010], Müger [2003b].

For a fusion category $\mathcal{C}$, the Drinfeld center $\mathcal{Z}(\mathcal{C})$ is a non-degenerate braided fusion category of Frobenius-Perron dimension $\operatorname{FPdim} \mathcal{Z}(\mathcal{C}) = (\operatorname{FPdim} \mathcal{C})^2$. Drinfeld centers of fusion categories are characterized as those non-degenerate braided fusion categories $\mathcal{Z}$ containing a Lagrangian algebra, that is, a separable commutative algebra A such that $\operatorname{Hom}_{\mathcal{C}}(\mathbf{1}, A) \cong k$ and $\operatorname{FPdim}(A)^2 = \operatorname{FPdim}(\mathcal{Z})$ Davydov, Müger, Nikshych, and Ostrik [2013].

Example 6.1. (*Group-theoretical braided fusion categories.*) Let G be a finite group and let ω be a normalized 3-cocycle on G. Let also $D^\omega G$-mod be the twisted quantum double Dijkgraaf, Pasquier, and Roche [1991]. Then there is an equivalence of braided fusion categories $D^\omega G\text{-mod} \cong \mathcal{Z}(\operatorname{Vect}_G^\omega)$. Suppose that $\mathcal{C}$ is a braided group-theoretical fusion category. Then there are equivalences of braided fusion categories $\mathcal{Z}(\mathcal{C}) \cong \mathcal{Z}(\operatorname{Vect}_G^\omega) \cong D^\omega G$-mod, for some finite group G and 3-cocycle ω. Thus, every braided group-theoretical fusion category can be embedded in $D^\omega G$-mod. A description of the fusion subcategories of $D^\omega G$-mod was given in Naidu, Nikshych, and Witherspoon [2009] in terms of subgroups of G and certain so-called *ω-bicharacters* on them.

Let G be a finite group. The fusion category $\operatorname{Rep} G$ is a symmetric fusion category with respect to the canonical braiding given by the flip of vector spaces. A braided fusion category $\mathcal{E}$ is called *Tannakian* if $\mathcal{E} \cong \operatorname{Rep} G$ for some finite group G as symmetric fusion categories. More generally, if $u \in G$ is a central element such that $u^2 = 1$, then the category $\operatorname{Rep}(G, u)$ of representations of G on finite dimensional super-vector spaces where u acts as the parity operator is a symmetric fusion category. The category $\operatorname{Rep}(\mathbb{Z}_2, u)$, where $1 \neq u \in \mathbb{Z}_2$ is denoted sVect.

The following result of Deligne is a crucial ingredient in the approach to the classification of braided fusion categories in the literature. Related results in a C^*-context are due to Doplicher and Roberts [1989].

Theorem 6.2. *Deligne [1990], Deligne [2002]. Let $\mathcal{C}$ be a symmetric fusion category. Then $\mathcal{C}$ is equivalent as a braided fusion category to the category* $\operatorname{Rep}(G, u)$ *for some finite group G and central element $u \in G$ such that $u^2 = 1$, which are, up to isomorphism, uniquely determined by* $\mathcal{C}$.

Let $\mathcal{C} \cong \operatorname{Rep}(G, u)$ be a symmetric fusion category. Then $\mathcal{C}$ is a $\mathbb{Z}_2$-extension of the Tannakian subcategory $\mathcal{E} = \operatorname{Rep}(G/(u))$. Thus if $\operatorname{FPdim} \mathcal{C} > 2$, then $\mathcal{C}$ contains a Tannakian subcategory, and a non-Tannakian symmetric fusion category of dimension 2 is equivalent to the category sVect.

Let $\mathcal{C}$ be a fusion category and suppose that $\mathcal{E} \cong \operatorname{Rep} G$ is a Tannakian subcategory of the center $\mathcal{Z}(\mathcal{C})$. The *de-equivariantization* of $\mathcal{C}$ with respect to G (or with respect to $\mathcal{E}$), denoted $\mathcal{C}_G$, is the category $\mathcal{C}_A$ of right A-modules in $\mathcal{C}$, where $A \in \mathcal{E}$ is the (commutative, separable) algebra corresponding to $k^G \in \operatorname{Rep} G$ under an equivalence of braided fusion categories $\operatorname{Rep} G \to \mathcal{E}$. This is a fusion category with tensor product $\otimes_A$ and unit object A. De-equivariantization and equivariantization are inverse processes. Indeed, the natural action by translations of G on A induces an action of G on $\mathcal{C}_G$ by tensor auto-equivalences such that $\mathcal{C} \cong (\mathcal{C}_G)^G$. See Bruguières [2000], Müger [2000], V. Drinfeld, Gelaki, Nikshych, and Ostrik [2010, Section 4].

6.1 Modular categories. A *premodular* category (Bruguières [2000]) is a braided fusion category endowed with a natural isomorphism $\theta : \mathrm{id}_{\mathcal{C}} \to \mathrm{id}_{\mathcal{C}}$, called a *ribbon structure*, satisfying

$$\theta_{X\otimes Y} = (\theta_X \otimes \theta_Y)c_{Y,X}c_{X,Y}, \quad \theta_X^* = \theta_{X^*}, \tag{6-1}$$

for all objects X, Y of $\mathcal{C}$. Equivalently, $\mathcal{C}$ is a braided fusion category endowed with a spherical structure. The ribbon structure of $\mathcal{C}$ allows to consider the *quantum trace* of an endomorphism $f : X \to X$ of an object X of $\mathcal{C}$ and in particular the *quantum dimension* of X defined as $\dim X = \operatorname{Tr}(\mathrm{id}_X)$.

Suppose $\mathcal{C}$ is a premodular category. Let X, Y be simple objects of $\mathcal{C}$ and let $S_{X,Y} \in k$ denote the quantum trace of the squared braiding $c_{Y,X}c_{X,Y} : X \otimes Y \to Y \otimes X$. The S*-matrix* of $\mathcal{C}$ is defined in the form $S = (S_{XY})_{X,Y\in\operatorname{Irr}(\mathcal{C})}$.

A premodular category $\mathcal{C}$ is called *modular* if its S-matrix is invertible. Equivalently, a premodular category $\mathcal{C}$ is modular if and only if it is non-degenerate. Every non-degenerate fusion category with integral Frobenius-Perron dimension is a modular category with its canonical spherical structure (c.f. Remark 3.4).

Let $\mathcal{C}$ be a modular category. The following relation, known as *Verlinde formula*, gives the fusion coefficients of $\mathcal{C}$ in terms of its S-matrix:

$$N_{XY}^Z = \frac{1}{\dim \mathcal{C}} \sum_{T\in\operatorname{Irr}(\mathcal{C})} \frac{S_{XT}\, S_{YT}\, S_{Z^*T}}{d_T},$$

for all $X, Y, Z \in \operatorname{Irr}(\mathcal{C})$, where d_T is the categorical dimension of the object T and $\dim \mathcal{C} = \sum_{T\in\operatorname{Irr}(\mathcal{C})} d_T^2$ is the categorical dimension of $\mathcal{C}$; c.f. Bakalov and Kirillov [2001].

Let $\mathcal{C}$ be a premodular category. A *modularization* of $\mathcal{C}$ is a dominant tensor functor $F : \mathcal{C} \to \mathcal{C}_0$ compatible with the braiding and the ribbon structures, where $\mathcal{C}_0$ is a modular category Bruguières [2000]. If such a modularization exists, then $\mathcal{C}$ is called *modularizable*. In Bruguières [ibid.], Müger [2000], Bruguières and Müger showed that a premodular category is modularizable if and only if its symmetric center $\mathcal{C}'$ is a Tannakian category. In this case, $\mathcal{C}' \cong \operatorname{Rep} G$ for some finite group G and $\mathcal{C}_0$ is the de-equivariantization $\mathcal{C}_G$. In this context, the group G acts on $\mathcal{C}_0$ by braided auto-equivalences and there is an equivalence of braided fusion categories $\mathcal{C} \cong \mathcal{C}_0^G$.

The following theorem was conjectured by Z. Wang. It implies the feasibility of the classification of modular fusion categories of a given rank:

Theorem 6.3. *Bruillard, Ng, E. C. Rowell, and Wang [2016b]. For every natural number r there is, up to equivalence, a finite number of modular categories of rank r.*

As pointed out by Etingof, the number of modular categories of rank r is, however, not bounded by any polynomial in r. Up to monoidal equivalence, the classification of modular categories has been achieved up to rank 5 Bruillard, Ng, E. C. Rowell, and Wang [2016a], E. Rowell, Stong, and Wang [2009].

6.2 Witt group of non-degenerate braided fusion categories. The (abelian) group $\mathcal{W}$ of Witt classes of non-degenerate braided fusion categories was introduced in Davydov, Müger, Nikshych, and Ostrik [2013]. Two non-degenerate braided fusion categories $\mathcal{C}_1$ and $\mathcal{C}_2$ are called *Witt equivalent* if there exist fusion categories $\mathcal{D}_1$ and $\mathcal{D}_2$ such that $\mathcal{C}_1 \boxtimes \mathcal{Z}(\mathcal{D}_1) \cong \mathcal{C}_2 \boxtimes \mathcal{Z}(\mathcal{D}_2)$ as braided tensor categories.

The Witt group $\mathcal{W}$ consists of equivalence classes of non-degenerate braided fusion categories under this equivalence relation with multiplication induced by Deligne's tensor product $\boxtimes$. The unit element is the class of the category Vect of finite-dimensional vector spaces over k and the inverse of the class of a non-degenerate braided fusion category $\mathcal{C}$ is the class of the *reverse* braided fusion category $\mathcal{C}^{rev}$ (this is the fusion category $\mathcal{C}$ with the *reversed* braiding $c^{rev}_{X,Y} = c^{-1}_{Y,X}$). The explicit determination of the relations in $\mathcal{W}$ is a relevant problem in connection with the classification of fusion categories.

Let $\mathcal{W}_{pt}$ and $\mathcal{W}_{Ising}$ denote the subgroups of Witt classes of pointed non-degenerate fusion categories and Ising braided categories, respectively. The subgroups $\mathcal{W}_{pt}$ and $\mathcal{W}_{Ising}$ were described in V. Drinfeld, Gelaki, Nikshych, and Ostrik [2010]. Let $\widetilde{\mathcal{W}}$ be the subgroup of $\mathcal{W}$ generated by the Witt classes of the modular categories $\mathcal{C}(\mathfrak{g}, l)$ of integrable highest weight modules of level $l \in \mathbb{Z}_+$ over the affinization of a simple finite-dimensional Lie algebra $\mathfrak{g}$ (see SubSection 2.3). It was shown in Davydov, Müger, Nikshych, and Ostrik [2013] that $\mathcal{W}_{pt}, \mathcal{W}_{Ising} \subseteq \widetilde{\mathcal{W}}$. The following open question was raised in Davydov, Müger, Nikshych, and Ostrik [ibid.] as a mathematical formulation of a conjecture stated by Moore and Seiberg [1989b]:

Question 6.4. Davydov, Müger, Nikshych, and Ostrik [2013]. Does $\widetilde{\mathcal{W}}$ coincide with be the subgroup $\mathcal{W}_{un}$ of Witt classes of pseudo-unitary non-degenerate braided fusion categories?

In relation with Question 5.7, we have:

Theorem 6.5. *Natale [2014]. The Witt class of a weakly group-theoretical non-degenerate braided fusion category belongs to the subgroup generated by $\mathcal{W}_{Ising}$ and $\mathcal{W}_{pt}$, whence also to $\widetilde{\mathcal{W}}$.*

6.3 Tannakian categories and braided G-crossed fusion categories. Let G be a finite group. A *G-crossed braided fusion category* is a fusion category $\mathfrak{D}$ endowed with a G-grading $\mathfrak{D} = \oplus_{g\in G} \mathfrak{A}_g$ and an action of G by tensor autoequivalences $\rho : \underline{G} \to \mathrm{Aut}_{\otimes}\ \mathfrak{D}$, such that $\rho^g(\mathfrak{D}_h) \subseteq \mathfrak{D}_{ghg^{-1}}$, for all $g, h \in G$, and a *G-braiding* $c : X \otimes Y \to \rho^g(Y) \otimes X$, $g \in G$, $X \in \mathfrak{D}_g$, $Y \in \mathfrak{D}$, subject to certain compatibility conditions. This notion was introduced by Turaev; see V. Turaev [2010].

The equivariantization $\mathfrak{D}^G$ of a G-crossed braided fusion category is a braided fusion category. The canonical embedding $\mathrm{Rep}\, G \to \mathfrak{D}^G$ of fusion categories is fact an embedding of braided fusion categories, with respect to the canonical braiding in $\mathrm{Rep}\, G$. The G-braiding on $\mathfrak{D}$ restricts to a braiding in the neutral component $\mathfrak{D}_1$ of the G-grading. Furthermore, the group G acts by restriction on $\mathfrak{D}_1$ and this action is by braided tensor autoequivalences. This makes the equivariantization $(\mathfrak{D}_1)^G$ into a braided fusion subcategory of $\mathfrak{D}^G$. This fusion subcategory coincides with the centralizer $\mathcal{E}'$ of the Tannakian subcategory $\mathcal{E} \cong \mathrm{Rep}\, G$ in $\mathfrak{D}^G$.

Conversely, if $\mathcal{C}$ is a braided fusion category containing a Tannakian subcategory $\mathcal{E} \cong \mathrm{Rep}\, G$, then the de-equivariantization $\mathcal{C}_G$ of $\mathcal{C}$ with respect to $\mathcal{E}$ is a braided G-crossed fusion category. Thus equivariantization defines a bijective correspondence between equivalence classes of braided fusion categories containing $\mathrm{Rep}\, G$ as a Tannakian subcategory and G-crossed braided fusion categories Jr [2001], Müger [2004].

Let $\mathfrak{D}$ be a G-crossed braided fusion category. The braided fusion category $\mathfrak{D}^G$ is non-degenerate if and only if the neutral component $\mathfrak{D}_1$ is non-degenerate and the G-grading of $\mathfrak{D}$ is faithful. If this is the case, then there is an equivalence of braided fusion categories $\mathcal{Z}(\mathfrak{D}) \cong \mathfrak{D}^G \boxtimes \mathfrak{D}_1^{rev}$ Davydov, Müger, Nikshych, and Ostrik [2013].

An important invariant of a braided fusion category is its *core*, introduced in V. Drinfeld, Gelaki, Nikshych, and Ostrik [2010]. As a braided fusion category, the core of a braided fusion category $\mathcal{C}$ is the neutral homogeneous component $\mathcal{C}_G^0$ of the de-equivariantization of $\mathcal{C}$ by a maximal Tannakian subcategory $\mathcal{E} \cong \mathrm{Rep}\, G$. The core of $\mathcal{C}$ is independent of $\mathcal{E}$. Furthermore, the core of a braided fusion category is *weakly anisotropic*, that is, it contains no Tannakian subcategories stable under all braided auto-equivalences. In addition, the core of $\mathcal{C}$ is non-degenerate if $\mathcal{C}$ is non-degenerate. A complete classification of pointed weakly anisotropic braided fusion categories has been proposed in V. Drinfeld, Gelaki, Nikshych, and Ostrik [ibid.].

A braided fusion category is weakly group-theoretical if and only if it can be obtained from a crossed braided fusion category with pointed neutral component, thus from a pointed category by means of suitable group extensions and equivariantizations:

Theorem 6.6. *Natale [2017]. Let $\mathcal{C}$ be a weakly group-theoretical braided fusion category. Then the core of $\mathcal{C}$ is equivalent to a Deligne tensor product $\mathfrak{B} \boxtimes \mathfrak{D}$, where $\mathfrak{D}$ is a* *pointed weakly anisotropic braided fusion category and* $\mathcal{B} \cong \mathrm{Vect}$ *or* $\mathcal{B}$ *is an Ising category. In particular, if* $\mathcal{C}$ *is integral, then its core is a pointed weakly anisotropic braided fusion category.*

The following theorem summarizes some known results related to Question 5.7 whose proofs rely on the existence of Tannakian subcategories and its relation with G-crossed braided fusion categories outlined above.

Theorem 6.7. *Let* $\mathcal{C}$ *be a braided fusion category. Then the following hold:*

(i) *Bruguières and Natale [2011]. If* $\operatorname{FPdim} \mathcal{C}$ *is an odd square-free integer, then* $\mathcal{C}$ *is equivalent to* $\operatorname{Rep} G$ *as a fusion category for some finite group* G.

(ii) *Natale [2014]. If* $\mathcal{C}$ *is non-degenerate and* $\operatorname{FPdim} \mathcal{C}$ *is a natural number less than* 1800*, or an odd natural number less than* 33075*, then* $\mathcal{C}$ *is weakly group-theoretical.*

(iii) *Natale and Pacheco Rodríguez [2016]. If* $\operatorname{FPdim} \mathcal{C} \in \mathbb{Z}$ *and the Frobenius-Perron dimensions of any simple object of* $\mathcal{C}$ *is a* p_i*-power, for some* $1 \leq i \leq r$*, where let* $p_1, \ldots, p_r$ *be prime numbers, then* $\mathcal{C}$ *is weakly group-theoretical. Moreover, it is solvable is either* $r \leq 2$*, or* $p_i > 7$*, for all* $i = 1, \ldots, r$.

(iv) *Natale [2017]. If* $\mathcal{C}$ *is integral and non-degenerate such that* $\operatorname{FPdim} X \leq 2$*, for every simple object* X*, then* $\mathcal{C}$ *is group-theoretical.*

(v) *Dong and Natale [2017]. If* $\mathcal{C}$ *is a non-degenerate and* $\operatorname{FPdim} \mathcal{C} = dq^n$*, where* $n \geq 0$*,* d *is a square-free natural number and* q *is an odd prime number, then* $\mathcal{C}$ *is group-theoretical.*

References

Henning Haahr Andersen and Jan Paradowski (1995). "Fusion categories arising from semisimple Lie algebras". *Comm. Math. Phys.* 169.3, pp. 563–588. MR: 1328736 (cit. on p. 196).

N. Andruskiewitsch and J. Devoto (1995). "Extensions of Hopf algebras". *Algebra i Analiz* 7.1, pp. 22–61. MR: 1334152 (cit. on p. 203).

Nicolás Andruskiewitsch (2002). "About finite dimensional Hopf algebras". In: *Quantum symmetries in theoretical physics and mathematics (Bariloche, 2000)*. Vol. 294. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 1–57. MR: 1907185 (cit. on p. 204).

Nicolás Andruskiewitsch and Sonia Natale (2006). "Tensor categories attached to double groupoids". *Adv. Math.* 200.2, pp. 539–583. MR: 2200856 (cit. on p. 196).

M. Asaeda and U. Haagerup (1999). "Exotic subfactors of finite depth with Jones indices $(5+\sqrt{13})/2$ and $(5+\sqrt{17})/2$". *Comm. Math. Phys.* 202.1, pp. 1–63. MR: 1686551 (cit. on p. 197).

Bojko Bakalov and Alexander Kirillov Jr. (2001). *Lectures on tensor categories and modular functors*. Vol. 21. University Lecture Series. American Mathematical Society, Providence, RI, pp. x+221. MR: 1797619 (cit. on pp. 192, 196, 212).

Stephen Bigelow, Emily Peters, Scott Morrison, and Noah Snyder (2012). "Constructing the extended Haagerup planar algebra". *Acta Math.* 209.1, pp. 29–82. MR: 2979509 (cit. on p. 197).

Gabriella Böhm, Florian Nill, and Kornél Szlachányi (1999). "Weak Hopf algebras. I. Integral theory and C^*-structure". *J. Algebra* 221.2, pp. 385–438. MR: 1726707 (cit. on p. 195).

Gabriella Bòhm and Korníl Szlachónyi (1996). "A coassociative C^*-quantum group with nonintegral dimensions". *Lett. Math. Phys.* 38.4, pp. 437–456. MR: 1421688 (cit. on p. 195).

Alain Bruguières (2000). "Catégories prémodulaires, modularisations et invariants des variétés de dimension 3". *Math. Ann.* 316.2, pp. 215–236. MR: 1741269 (cit. on p. 212).

Alain Bruguières, Steve Lack, and Alexis Virelizier (2011). "Hopf monads on monoidal categories". *Adv. Math.* 227.2, pp. 745–800. MR: 2793022 (cit. on pp. 194, 196).

Alain Bruguières and Sonia Natale (2011). "Exact sequences of tensor categories". *Int. Math. Res. Not. IMRN* 24, pp. 5644–5705. MR: 2863377 (cit. on pp. 194, 202, 203, 215).

– (2014). "Central exact sequences of tensor categories, equivariantization and applications". *J. Math. Soc. Japan* 66.1, pp. 257–287. MR: 3161401 (cit. on pp. 203, 208).

Alain Bruguières and Alexis Virelizier (2007). "Hopf monads". *Adv. Math.* 215.2, pp. 679–733. MR: 2355605 (cit. on p. 194).

Paul Bruillard, Siu-Hung Ng, Eric C. Rowell, and Zhenghan Wang (2016a). "On classification of modular categories by rank". *Int. Math. Res. Not. IMRN* 24, pp. 7546–7588. arXiv: 1507.05139. MR: 3632091 (cit. on p. 213).

– (2016b). "Rank-finiteness for modular categories". *J. Amer. Math. Soc.* 29.3, pp. 857–881. MR: 3486174 (cit. on pp. 198, 213).

Frank Calegari, Scott Morrison, and Noah Snyder (2011). "Cyclotomic integers, fusion categories, and subfactors". *Comm. Math. Phys.* 303.3, pp. 845–896. MR: 2786219 (cit. on p. 199).

Alexei Davydov, Michael Müger, Dmitri Nikshych, and Victor Ostrik (2013). "The Witt group of non-degenerate braided fusion categories". *J. Reine Angew. Math.* 677, pp. 135–177. MR: 3039775 (cit. on pp. 211, 213, 214).

P. Deligne (1990). "Catégories tannakiennes". In: *The Grothendieck Festschrift, Vol. II*. Vol. 87. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 111–195. MR: 1106898 (cit. on pp. 204, 211).

– (2002). "Catégories tensorielles". *Mosc. Math. J.* 2.2. Dedicated to Yuri I. Manin on the occasion of his 65th birthday, pp. 227–248. MR: 1944506 (cit. on p. 211).

R. Dijkgraaf, V. Pasquier, and P. Roche (1991). "Quasi-quantum groups related to orbifold models". In: *Modern quantum field theory (Bombay, 1990)*. World Sci. Publ., River Edge, NJ, pp. 375–383. MR: 1237434 (cit. on pp. 205, 211).

Jingcheng Dong and Sonia Natale (Nov. 2017). "On the classification of almost square-free modular categories" (cit. on p. 215).

Sergio Doplicher and John E. Roberts (1989). "A new duality theory for compact groups". *Invent. Math.* 98.1, pp. 157–218. MR: 1010160 (cit. on p. 211).

V. G. Drinfeld (1989a). "Almost cocommutative Hopf algebras". *Algebra i Analiz* 1.2, pp. 30–46. MR: 1025154 (cit. on p. 194).

– (1989b). "Quasi-Hopf algebras". *Algebra i Analiz* 1.6, pp. 114–148. MR: 1047964 (cit. on p. 193).

Vladimir Drinfeld, Shlomo Gelaki, Dmitri Nikshych, and Victor Ostrik (Apr. 2007). "Group-theoretical properties of nilpotent modular categories". arXiv: 0704.0195 (cit. on p. 206).

– (2010). "On braided fusion categories. I". *Selecta Math. (N.S.)* 16.1, pp. 1–119. MR: 2609644 (cit. on pp. 210, 212–214).

Melisa Escañuela González and Sonia Natale (2017). "On fusion rules and solvability of a fusion category". *J. Group Theory* 20.1, pp. 133–167. MR: 3592609 (cit. on p. 210).

Pavel Etingof and Shlomo Gelaki (2001). "Isocategorical groups". *Internat. Math. Res. Notices* 2, pp. 59–76. MR: 1810480 (cit. on p. 195).

– (2017). "Exact sequences of tensor categories with respect to a module category". *Adv. Math.* 308, pp. 1187–1208. MR: 3600085 (cit. on pp. 204, 205).

Pavel Etingof, Shlomo Gelaki, Dmitri Nikshych, and Victor Ostrik (2015). *Tensor categories*. Vol. 205. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. xvi+343. MR: 3242743 (cit. on pp. 192, 198).

Pavel Etingof, Shlomo Gelaki, and Viktor Ostrik (2004). "Classification of fusion categories of dimension pq". *Int. Math. Res. Not.* 57, pp. 3041–3056. MR: 2098028 (cit. on p. 206).

Pavel Etingof, Dmitri Nikshych, and Victor Ostrik (2010). "Fusion categories and homotopy theory". *Quantum Topol.* 1.3. With an appendix by Ehud Meir, pp. 209–273. MR: 2677836 (cit. on p. 207).

– (2011). "Weakly group-theoretical and solvable fusion categories". *Adv. Math.* 226.1, pp. 176–205. MR: 2735754 (cit. on pp. 202, 206, 208–210).

Pavel Etingof, Dmitri Nikshych, and Viktor Ostrik (2005). "On fusion categories". *Ann. of Math. (2)* 162.2, pp. 581–642. MR: 2183279 (cit. on pp. 191, 197–200, 202, 206, 210).

David E. Evans and Terry Gannon (2014). "Near-group fusion categories and their doubles". *Adv. Math.* 255, pp. 586–640. MR: 3167494 (cit. on p. 200).

M. Finkelberg (1996). "An equivalence of fusion categories". *Geom. Funct. Anal.* 6.2, 249–267. (Erratum: Geom. Funct. Anal. **23**, 810–811 (2013)). MR: 1384612 (cit. on p. 197).

César Galindo and Sonia Natale (2007). "Simple Hopf algebras and deformations of finite groups". *Math. Res. Lett.* 14.6, pp. 943–954. MR: 2357466 (cit. on pp. 203, 204).

Shlomo Gelaki and Dmitri Nikshych (2008). "Nilpotent fusion categories". *Adv. Math.* 217.3, pp. 1053–1071. MR: 2383894 (cit. on pp. 199, 207).

Pinhas Grossman, Masaki Izumi, and Noah Snyder (Jan. 2015). "The Asaeda-Haagerup fusion categories". arXiv: 1501.07324 (cit. on p. 197).

Uffe Haagerup (1994). "Principal graphs of subfactors in the index range $4 < [M : N] < 3 + \sqrt{2}$". In: *Subfactors (Kyuzeso, 1993)*. World Sci. Publ., River Edge, NJ, pp. 1–38. MR: 1317352 (cit. on p. 197).

Frank Hausser and Florian Nill (1999). "Doubles of quasi-quantum groups". *Comm. Math. Phys.* 199.3, pp. 547–589. MR: 1669685 (cit. on p. 194).

Takahiro Hayashi (2000). "A brief introduction to face algebras". In: *New trends in Hopf algebra theory (La Falda, 1999)*. Vol. 267. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 161–176. MR: 1800711 (cit. on p. 196).

Yi-Zhi Huang (2008). "Rigidity and modularity of vertex tensor categories". *Communications in Contemporary Mathematics* 10.Suppl. 1, pp. 871–911. MR: 2468370 (cit. on p. 197).

Yi-Zhi Huang and James Lepowsky (1999). "Intertwining operator algebras and vertex tensor categories for affine Lie algebras". *Duke Math. J.* 99.1, pp. 113–134. MR: 1700743 (cit. on p. 197).

Masaki Izumi (2001). "The structure of sectors associated with Longo-Rehren inclusions. II. Examples". *Rev. Math. Phys.* 13.5, pp. 603–674. MR: 1832764 (cit. on p. 197).

V. F. R. Jones (1983). "Index for subfactors". *Invent. Math.* 72.1, pp. 1–25. MR: 696688 (cit. on p. 197).

Vaughan F. R. Jones, Scott Morrison, and Noah Snyder (2014). "The classification of subfactors of index at most 5". *Bull. Amer. Math. Soc. (N.S.)* 51.2, pp. 277–327. MR: 3166042 (cit. on p. 197).

André Joyal and Ross Street (1993). "Braided tensor categories". *Adv. Math.* 102.1, pp. 20–78. MR: 1250465 (cit. on p. 193).

Alexander Kirillov Jr (Oct. 2001). "Modular categories and orbifold models II". arXiv: math/0110221 (cit. on p. 214).

G. I. Kac (1962). "Finite ring groups". *Dokl. Akad. Nauk SSSR* 147, pp. 21–24. MR: 0145366 (cit. on p. 206).

Christian Kassel (1995). *Quantum groups*. Vol. 155. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xii+531. MR: 1321145 (cit. on p. 192).

George Lusztig (1993). *Introduction to quantum groups*. Vol. 110. Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, pp. xii+341. MR: 1227098 (cit. on p. 196).

Saunders Mac Lane (1998). *Categories for the working mathematician*. Second. Vol. 5. Graduate Texts in Mathematics. Springer, New York, pp. xii+314. MR: 1712872 (cit. on p. 194).

S. Majid (1998). "Quantum double for quasi-Hopf algebras". *Lett. Math. Phys.* 45.1, pp. 1–9. MR: 1631648 (cit. on pp. 194, 205).

Shahn Majid (1995). *Foundations of quantum group theory*. Cambridge University Press, Cambridge, pp. x+607. MR: 1381692 (cit. on p. 192).

Akira Masuoka (2002). "Hopf algebra extensions and cohomology". In: *New directions in Hopf algebras*. Vol. 43. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, pp. 167–209. MR: 1913439 (cit. on p. 206).

Gregory Moore and Nathan Seiberg (1989a). "Classical and quantum conformal field theory". *Comm. Math. Phys.* 123.2, pp. 177–254. MR: 1002038 (cit. on p. 200).

– (1989b). "Taming the conformal zoo". *Phys. Lett. B* 220.3, pp. 422–430. MR: 992362 (cit. on p. 213).

Scott Morrison and Noah Snyder (2012). "Non-cyclotomic fusion categories". *Trans. Amer. Math. Soc.* 364.9, pp. 4713–4733. MR: 2922607 (cit. on p. 197).

Michael Müger (2000). "Galois theory for braided tensor categories and the modular closure". *Adv. Math.* 150.2, pp. 151–201. MR: 1749250 (cit. on p. 212).

– (2003a). "From subfactors to categories and topology. I. Frobenius algebras in and Morita equivalence of tensor categories". *J. Pure Appl. Algebra* 180.1-2, pp. 81–157. MR: 1966524 (cit. on pp. 200, 202).

– (2003b). "On the structure of modular categories". *Proc. London Math. Soc. (3)* 87.2, pp. 291–308. MR: 1990929 (cit. on p. 210).

– (2004). "Galois extensions of braided tensor categories and braided crossed G-categories". *J. Algebra* 277.1, pp. 256–281. MR: 2059630 (cit. on p. 214).

Deepak Naidu, Dmitri Nikshych, and Sarah Witherspoon (2009). "Fusion subcategories of representation categories of twisted quantum doubles of finite groups". *Int. Math. Res. Not. IMRN* 22, pp. 4183–4219. MR: 2552301 (cit. on p. 211).

Sonia Natale (2007). "Semisolvability of semisimple Hopf algebras of low dimension". *Mem. Amer. Math. Soc.* 186.874, pp. viii+123. MR: 2294999 (cit. on p. 204).

– (2010). "Semisimple Hopf algebras of dimension 60". *J. Algebra* 324.11, pp. 3017–3034. MR: 2732985 (cit. on p. 204).

– (2014). "On weakly group-theoretical non-degenerate braided fusion categories". *J. Noncommut. Geom.* 8.4, pp. 1043–1060. MR: 3310938 (cit. on pp. 210, 213, 215).

– (2015). "Jordan-Hölder theorem for finite dimensional Hopf algebras". *Proc. Amer. Math. Soc.* 143.12, pp. 5195–5211. MR: 3411137 (cit. on p. 204).

Sonia Natale (2016a). "A Jordan-Hölder theorem for weakly group-theoretical fusion categories". *Math. Z.* 283.1-2, pp. 367–379. MR: 3489071 (cit. on p. 210).

– (2016b). "Crossed actions of matched pairs of groups on tensor categories". *Tohoku Math. J. (2)* 68.3, pp. 377–405. MR: 3550925 (cit. on p. 208).

– (Apr. 2017). "The core of a weakly group-theoretical braided fusion category". arXiv: 1704.03523. MR: 3770935 (cit. on pp. 214, 215).

Sonia Natale and Edwin Pacheco Rodríguez (2016). "Graphs attached to simple Frobenius–Perron dimensions of an integral fusion category". *Monatsh. Math.* 179.4, pp. 615–649. MR: 3474884 (cit. on pp. 200, 215).

Dmitri Nikshych (2008). "Non-group-theoretical semisimple Hopf algebras from group actions on fusion categories". *Selecta Math. (N.S.)* 14.1, pp. 145–161. MR: 2480712 (cit. on pp. 206–208).

Dmitri Nikshych and Leonid Vainerman (2002). "Finite quantum groupoids and their applications". In: *New directions in Hopf algebras*. Vol. 43. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, pp. 211–262. MR: 1913440 (cit. on p. 195).

Victor Ostrik (2003a). "Module categories, weak Hopf algebras and modular invariants". *Transform. Groups* 8.2, pp. 177–206. MR: 1976459 (cit. on p. 196).

– (2015). "Pivotal fusion categories of rank 3". *Mosc. Math. J.* 15.2, pp. 373–396, 405. MR: 3427429 (cit. on p. 199).

Viktor Ostrik (2003b). "Fusion categories of rank 2". *Math. Res. Lett.* 10.2-3, pp. 177–183. MR: 1981895 (cit. on pp. 198–200).

– (2003c). "Module categories over the Drinfeld double of a finite group". *Int. Math. Res. Not.* 27, pp. 1507–1520. MR: 1976233 (cit. on p. 205).

Emily Peters (2010). "A planar algebra construction of the Haagerup subfactor". *Internat. J. Math.* 21.8, pp. 987–1045. MR: 2679382 (cit. on p. 197).

Eric Rowell, Richard Stong, and Zhenghan Wang (2009). "On classification of modular tensor categories". *Comm. Math. Phys.* 292.2, pp. 343–389. MR: 2544735 (cit. on p. 213).

Stephen F. Sawin (2006). "Quantum groups at roots of unity and modularity". *J. Knot Theory Ramifications* 15.10, pp. 1245–1277. MR: 2286123 (cit. on p. 196).

Jacob Siehler (2003). "Near-group categories". *Algebr. Geom. Topol.* 3, pp. 719–775. MR: 1997336 (cit. on p. 200).

Daisuke Tambara (2001). "Invariants and semi-direct products for finite group actions on tensor categories". *J. Math. Soc. Japan* 53.2, pp. 429–456. MR: 1815142 (cit. on pp. 208, 209).

Daisuke Tambara and Shigeru Yamagami (1998). "Tensor categories with fusion rules of self-duality for finite abelian groups". *J. Algebra* 209.2, pp. 692–707. MR: 1659954 (cit. on pp. 198, 201).

V. G. Turaev (1994). *Quantum invariants of knots and 3-manifolds*. Vol. 18. De Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin, pp. x+588. MR: 1292673 (cit. on pp. 192, 196).

Vladimir Turaev (2010). *Homotopy quantum field theory*. Vol. 10. EMS Tracts in Mathematics. Appendix 5 by Michael Müger and Appendices 6 and 7 by Alexis Virelizier. European Mathematical Society (EMS), Zürich, pp. xiv+276. MR: 2674592 (cit. on p. 214).

Received 2017-12-02.

Sonia Natale
Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba
CIEM – CONICET
(5000) Ciudad Universitaria
Córdoba
Argentina
natale@famaf.unc.edu.ar

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (219–240)

ON GROTHENDIECK–SERRE CONJECTURE CONCERNING PRINCIPAL BUNDLES

Ivan Panin

Abstract

Let R be a regular local ring. Let $\mathbf{G}$ be a reductive group scheme over R. A well-known conjecture due to Grothendieck and Serre assertes that a principal $\mathbf{G}$-bundle over R is trivial, if it is trivial over the fraction field of R. In other words, if K is the fraction field of R, then the map of non-abelian cohomology pointed sets

$$\mathrm{H}^1_{\text{ét}}(R, \mathbf{G}) \to \mathrm{H}^1_{\text{ét}}(K, \mathbf{G}),$$

induced by the inclusion of R into K, has a trivial kernel. *The conjecture is solved in positive for all regular local rings contaning a field.* More precisely, if the ring R contains an infinite field, then this conjecture is proved in a joint paper due to R. Fedorov and I. Panin published in 2015 in Publications l'IHES. If the ring R contains a finite field, then this conjecture is proved in 2015 in a preprint due to I. Panin which can be found on preprint server Linear Algebraic Groups and Related Structures. A more structured exposition can be found in Panin's preprint of the year 2017 on arXiv.org.

This and other results concerning the conjecture are discussed in the present paper. We illustrate the exposition by many interesting examples. We begin with couple results for complex algebraic varieties and develop the exposition step by step to its full generality.

Contents

The author acknowledges support of the Program of the Presidium of the Russian Academy of Sciences No. 01 Fundamental Mathematics and its Applications (grant PRAS-18-01).
MSC2010: primary 20G15; secondary 20G35, 20G10.
Keywords: Linear algebraic groups, principal bundles, affine algebraic varieties.

1 Introduction

The conjecture was stated by J.-P. Serre in 1958 in so called constant case and by A. Grothendieck in 1968 in the general case. In this introduction we give couple results motivating the conjecture in the constant case. To do that recall some notation. Let X be an affine complex algebraic variety, smooth and irreducible. Let $\mathbb{C}[X]$ be the ring of regular functions on X and $f \in \mathbb{C}[X]$ be a non-zero function. Let $X_f := \{x \in X : f(x) \neq 0\}$. This open subset is called the principal open subset of X corresponding to the function f. This open subset X_f is itself is an affine algebraic variety and its ring of regular functions $\mathbb{C}[X_f]$ is the localization $\mathbb{C}[X]_f$ of the ring $\mathbb{C}[X]$ with respect to the element f. If A is a $\mathbb{C}[X]$-algebra, then we write A_f for the localization of A with respect to $f \in \mathbb{C}[X]$. We are ready now to formulate first result, which is due to Serre [1958].

Let A be a $\mathbb{C}[X]$-algebra, which is a free finitely generated $\mathbb{C}[X]$-module of rank n. Suppose that A is isomorphic to the matrix algebra $M_r(\mathbb{C}[X])$ locally for the complex topology on X. Suppose further that for a non-zero function $f \in \mathbb{C}[X]$ the $\mathbb{C}[X_f]$-algebras A_f and $M_r(\mathbb{C}[X_f])$ are isomorphic. Then for any point $x \in X$ there is a regular function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the $\mathbb{C}[X_g]$-algebras A_g and $M_r(\mathbb{C}[X_g])$ are isomorphic. In the other words, the $\mathbb{C}[X]$-algebras A and $M_r(\mathbb{C}[X])$ are isomorphic locally for the Zarisky topology on X.

Let us point out that the $\mathbb{C}[X]$-algebras A and $M_r(\mathbb{C}[X])$ are isomorphic locally for the complex topology on X by the assumption of the theorem. The theorem states that these $\mathbb{C}[X]$-algebras are isomorphic *locally for the Zarisky topology on* X provided that they are isomorphic over *a non-empty Zarisky open subset of* X.

It is easy to give many examples of $\mathbb{C}[X]$-algebras A, which are isomorphic to the matrix algebra $M_2(\mathbb{C}[X])$ locally for the complex topology, but which are not isomorphic to the matrix algebra $M_2(\mathbb{C}[X])$ locally for the Zarisky topology. These algebras can be found for instance among generalized quaternion algebras. Second result illustrating the conjecture is due to Ojanguren [1980].

Let X and $\mathbb{C}[X]$ be as above and let $a_i, b_i \in \mathbb{C}[X]$ be invertible functions on X, where $i \in \{1, ..., r\}$. Consider two quadratic spaces $P := \Sigma_{i=1}^r a_i T_i^2$ and $Q := \Sigma_{i=1}^r b_i T_i^2$ over $\mathbb{C}[X]$. Suppose for a non-zero function $f \in \mathbb{C}[X]$ these quadratic spaces are isomorphic over the ring $\mathbb{C}[X_f]$. Then the quadratic spaces P and Q are isomorphic locally for the Zarisky topology on X. In other words, for any point $x \in X$ there is a regular function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and quadratic spaces P and Q are isomorphic as quadratic spaces over $\mathbb{C}[X]_g$.

A very partial case of the above Serre's result can be formulated as follows. Let a and b two invertible regular functions on X. Consider so called generalized quaternion $\mathbb{C}[X]$-algebra A given by two generators u, v subjecting to defining relations $u^2 = a$, $v^2 = b$, $uv = -vu$. Suppose the $\mathbb{C}(X)$-algebra $A \otimes_{\mathbb{C}[X]} \mathbb{C}(X)$ is isomorphic to the matrix algebra $M_2(\mathbb{C}(X))$. Then the $\mathbb{C}[X]$-algebra A is isomorphic to the matrix algebra $M_2(\mathbb{C}[X])$ locally for the Zarisky topology on X.

The indicated results can be restated in terms of principal bundles for groups PGL_r, SO_r and PGL_2 respectively. It is pretty clear now that one can try to state a rather general theorem in terms of principal homogeneous spaces. That will be done in the next section.

2 Principal G-bundles

Recall some notion. Let $\mathbf{G}$ be a linear complex algebraic group, that is a closed subgroup of the general linear group $\mathrm{GL}_n(\mathbb{C})$ with respect to the Zarisky topology. Let X be as above. Let $(E, \nu : \mathbf{G} \times E \to E)$ be a pair such that E is a complex algebraic variety together with a regular map $p : E \to X$ and ν is a $\mathbf{G}$-action on E respecting to the map p, that is $p(\nu(g, e)) = p(e)$ for any $e \in E$ and $g \in \mathbf{G}$. We will write ge for $\nu(g, e)$.

A principal $\mathbf{G}$*-bundle over* X is a pair $(E, \nu : \mathbf{G} \times E \to E)$ above such that
1) the regular map $\mathbf{G} \times E \to E \times_X E$ taking (g, e) to (ge, e) is an isomorphism of algebraic varieties;
2) for any point $x \in X$ there are a neighborhood V of x in the complex topology on X and an isomorphism of complex holomorphic varieties $\varphi : E|_V := p^{-1}(V) \to \mathbf{G} \times V$ such that φ respects to the projection on V and φ respects to obvious $\mathbf{G}$-actions on both sides.

A morphism between two principal $\mathbf{G}$-bundles (E_1, ν_1) and (E_2, ν_2) is a morphism $\psi : E_1 \to E_2$ which respects as to the projections on the base X, so to the $\mathbf{G}$-actions. A trivialized $\mathbf{G}$-bundle is the $\mathbf{G}$-bundle $(\mathbf{G} \times X, \mu)$, where $\mu(g'(g, x)) = (g'g)x$. A trivial $\mathbf{G}$-bundle is a $\mathbf{G}$-bundle isomorphic to the trivialized one. Clearly, a $\mathbf{G}$-bundle (E, ν) over X is trivial if there is a section $s : X \to E$ of the projection $p : E \to X$. For a principal $\mathbf{G}$-bundle (E, ν) over X we often will write just E skipping ν from the notation. We will write often a $\mathbf{G}$-bundle for a principal $\mathbf{G}$-bundle.

Many examples of principal **G**-bundles are obtained by the following simple construction. Consider a closed embedding of algebraic groups $\mathbf{G} \subset \mathrm{GL}_n(\mathbb{C})$ and set $X = \mathbf{G}\backslash\mathrm{GL}_n(\mathbb{C})$ (the orbit variety of right cosets with respect to **G**). Then the pair $(\mathrm{GL}_n(\mathbb{C}), \nu : \mathbf{G} \times \mathrm{GL}_n(\mathbb{C}) \to GL_n(\mathbb{C}))$, where ν takes (g, h) to gh is a principal **G**-bundle over X. The fibres of the projection $p : \mathrm{GL}_n(\mathbb{C}) \to X$ are right cosets of $\mathrm{GL}_n(\mathbb{C})$ with respect to the subgroup **G**. This principal **G**-bundle is very often not trivial locally for the Zarisky topology. For instance, if **G** is the special orthogonal group $\mathrm{SO}_n(\mathbb{C}) \subset \mathrm{GL}_n(\mathbb{C})$, then the above principal G-bundle *is not trivial locally for the Zarisky topology.*

A principal **G**-bundle E over X is not necessary trivial locally for the Zarisky topology on X, *however it is always trivial locally for the étale topology on* X. The latter means the following: there is a regular map $\pi : X' \to X$ of smooth algebraic varieties and a regular map $s' : X' \to E$ such that $p \circ s' = \pi$, π is surjective and for any point $x' \in X'$ the induced map of tangent spaces $T_{X',x'} \to T_{X,\pi(u)}$ is an isomorphism. In the other words, $\pi : X' \to X$ is a surjective étale regular map and the principal **G**-bundle $X' \times_X E$ over X' is trivial. Indeed, the regular map $(id_{X'}, s') : X' \to X' \times_X E$ is a section of the projection $pr_{X'}$.

We are ready now to state a very general result concerning principal **G**-bundles and extending the results from the introduction.

Theorem 2.1. Let **G** be a simple (or a semi-simple, or even a reductive) complex algebraic group. Let X be an affine complex algebraic variety, smooth and irreducible and let E_1, E_2 be two principal **G**-bundles over X. Suppose there is a non-zero regular function $f \in \mathbb{C}[X]$ such that the principal **G**-bundles $E_1|_{X_f}$ and $E_2|_{X_f}$ are isomorphic over X_f. Then the principal **G**-bundles E_1 and E_2 *are isomorphic locally for the Zarisky topology on* X.

In other words, for any point $x \in X$ there is a regular function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the principal **G**-bundles $E_1|_{X_g}$ and $E_2|_{X_g}$ are isomorphic over X_g.

Remark 2.2. Let E_2 be a trivial principal **G**-bundle, then this theorem states the following. If E_1 is trivial over a non-empty Zarisky open subset of X, then E_1 *is trivial locally for the Zarisky topology on* X.

If E_2 is a trivial , then this theorem is due to Colliot-Thélène and Ojanguren [1992].

The general case of the theorem is due to R. Fedorov and the speaker Fedorov and I. Panin [2015].

Remark 2.3. The case when E_2 is trivial corresponds to the so called ”constant” case of the conjecture. Indeed, in this case one principal **G**-bundle $E = E_1$ is given and the result is a theorem about that principal **G**-bundle. The general case of the theorem can not be reduced to an equivalent statement concerning a principal **G** for the group **G**.

Let us give now few examples illustrating Theorem 2.1. Those examples are partial cases of Theorem 2.1 for the projective linear groups PGL_2,PGL_n, for the exceptional group G_2 and for the special projective orthogonal PSO_{2n} and for the projective simplectic groups PSp_{2n} respectively.
(1) Let A_1, A_2 be two generalized quaternion $\mathbb{C}[X]$-algebras corresponding to pairs a_1, b_1 and a_2, b_2 respectively. Suppose for a non-zero function $f \in \mathbb{C}[X]$ the $\mathbb{C}[X_f]$-algebras $(A_1)_f$ and $(A_2)_f$ are isomorphic. Then the $\mathbb{C}[X]$-algebras A_1 and A_2 are isomorphic locally for the Zarisky topology on X.
(2) Let A_1 and A_2 be two algebras as in the Serre's theorem above. *They are called Azumaya* $\mathbb{C}[X]$*-algebras.* Suppose for a non-zero function $f \in \mathbb{C}[X]$ the $\mathbb{C}[X_f]$-algebras $(A_1)_f$ and $(A_2)_f$ are isomorphic. Then the $\mathbb{C}[X]$-algebras A_1 and A_2 are isomorphic locally for the Zarisky topology on X.
(3) Let a_1, b_1, c_1 be invertible functions in $\mathbb{C}[X]$ and let $\mathbb{O}(a_1, b_1, c_1)$ be a generalized octonion $\mathbb{C}[X]$-algebra. That is as a $\mathbb{C}[X]$-module $\mathbb{O}(a_1, b_1, c_1)$ is a free of rank 8 with a free basis $1, e_1, e_2, ..., e_7$. And the multiplication table is as follows: $e_1^2 = a_1, e_2^2 = b_1, e_3^2 = c_1$

$$e_4 = e_1e_2 = -e_2e_1, e_5 = e_2e_3 = -e_3e_2, e_6 = e_3e_4 = -e_4e_3, e_7 = e_4e_5 = -e_5e_4.$$

Let a_2, b_2, c_2 be invertible functions in $\mathbb{C}[X]$ and let $\mathbb{O}(a_2, b_2, c_2)$ be one more generalized octonion $\mathbb{C}[X]$-algebra. Suppose for a non-zero function $f \in \mathbb{C}[X]$ the $\mathbb{C}[X_f]$-algebras $\mathbb{O}(a_1, b_1, c_1)_f$ and $\mathbb{O}(a_2, b_2, c_2)_f$ are isomorphic. Then the $\mathbb{C}[X]$-algebras $\mathbb{O}(a_1, b_1, c_1)$ and $\mathbb{O}(a_2, b_2, c_2)$ are isomorphic locally for the Zarisky topology on X.
(4) Let (A_1, σ_1) be an Azumaya $\mathbb{C}[X]$-algebra with involution. That is A_1 is an Azumaya algebra and $\sigma_1 : A_1 \to A_1^{op}$ is an $\mathbb{C}[X]$-algebras isomorphism, where A_1^{op} is the opposite $\mathbb{C}[X]$-algebra. Let (A_2, σ_2) be one more Azumaya $\mathbb{C}[X]$-algebra with involution. Suppose for a non-zero function $f \in \mathbb{C}[X]$ the $\mathbb{C}[X_f]$-algebras $(A_1, \sigma_1)_f$ and $(A_2, \sigma_2)_f$ are isomorphic as $\mathbb{C}[X_f]$-algebras with involutions. That is there is a $\mathbb{C}[X_f]$-algebras isomorphism $\varphi : (A_1)_f \to (A_2)_f$ with $\sigma_2 \circ \varphi = \varphi \circ \sigma_1$. Then the $\mathbb{C}[X]$-algebras A_1 and A_2 are isomorphic locally for the Zarisky topology on X as $\mathbb{C}[X]$-algebras with involutions.

3 Non-constant case of the conjecture for complex algebraic varieties

We begin with few examples illustrating the conjecture in the non-constant case. Let X be as above an affine complex algebraic variety, smooth and irreducible.
1) Let a, b be invertible functions in $\mathbb{C}[X]$. Consider an equation

$$T_1^2 - aT_2^2 = b \tag{1}$$

If this equation has a solution over the field $\mathbb{C}(X)$ of rational functions on X, then for any point $x \in X$ there is a function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the Equation (1) has a solution in $\mathbb{C}[X_g]$.
2) Let $t^n + a_{n-1}t^{n-1} + ... + a_0 = F(t) \in \mathbb{C}[X]$ with $a_i \in \mathbb{C}[X]$ be a monic polinomial such that for any point $x \in X$ the polinomial $F_x(t) = t^n + a_{n-1}(x)t^{n-1} + ... + a_0(x) \in \mathbb{C}[t]$ has no multiple roots. Let $X' \subset X \times \mathbb{A}^1$ be a closed subvariety defined by the equation $\{F = 0\}$. The regular function ring $\mathbb{C}[X']$ is the factor ring $\mathbb{C}[X][t]/(F)$. It is a free rank n module over $\mathbb{C}[X]$. Therefore there is a norm map $N_{\mathbb{C}[X']/\mathbb{C}[X]} : \mathbb{C}[X']^\times \to \mathbb{C}[X]^\times$, which takes an element α to the element $det(m_\alpha)$, where $m_\alpha : \mathbb{C}[X'] \to \mathbb{C}[X']$ is a multiplication by α. Let $a \in \mathbb{C}[X]^\times$ be a unit. Suppose the equation

$$N_{\mathbb{C}(X')/\mathbb{C}(X)}(\alpha) = a \tag{2}$$

has a solution in $\mathbb{C}(X')$, then for any point $x \in X$ there is a function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the equation $N_{\mathbb{C}[X'_g]/\mathbb{C}[X_g]}(\alpha) = a$ has a solution in $\mathbb{C}[X']_g$.
3) Let $a, b, c \in \mathbb{C}[X]$ be invertible functions. Consider an equation

$$T_1^2 - aT_2^2 - bT_3^2 + abT_4^2 = c \tag{3}$$

Suppose this equation has a solution over the field $\mathbb{C}(X)$. Then for any point $x \in X$ there is a function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the Equation (3) has a solution in $\mathbb{C}[X_g]$.
4) Let A be an Azumaya $\mathbb{C}[X]$-algebra of rank n from the introduction. Let $Nrd : A \to \mathbb{C}[X]$ be the reduced norm map. It is a map such that for any point $x \in X$ the map $A/m_xA = M_r(\mathbb{C}) \to \mathbb{C}$ is the determinant. Let $a \in \mathbb{C}[X]^\times$. Suppose the equation

$$Nrd(\alpha) = a \tag{4}$$

has a solution in $A \otimes_{\mathbb{C}[X]} \mathbb{C}(X)$. Then for any point $x \in X$ there is a function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the Equation (4) has a solution in A_g.
5) Let $Q = \Sigma_{i=1}^r b_i T_i^2$ be the quadratic space over $\mathbb{C}[X]$ from the introduction. Let $Q_{\mathbb{C}(X)}$ be the same quadratic space regarded over the field $\mathbb{C}(X)$. Let $a \in \mathbb{C}[X]^\times$ be a unite. Suppose a is a spinor norm over the field $\mathbb{C}(X)$, that is $a = Q(v_1) \cdot ... \cdot Q(v_{2n})$ for some vectors $v_1, ..., v_{2n} \in \mathbb{C}(X)^r$. Then a is a spinor norm locally for the Zarisky topology on X. In the other words, for any point x in X there is a function $g \in \mathbb{C}[X]$ not vanishing at x and there is a non-negative integer s and there are vectors $w_1, \dots, w_{2s} \in \mathbb{C}[X_g]^r$ such that $a = Q(w_1) \cdot \ldots \cdot Q(w_{2s})$.
6) Let (A, σ) be an Azumaya $\mathbb{C}[X]$-algebra with involution and let the involution be orthogonal, that is σ corresponds to a quadratic space locally for the complex topology on X. We use the terminology of Definition 3.1 here. Let $\mathrm{SO}_{A,\sigma} \subset \mathrm{GL}_{1,A}$ be the special orthogonal X-group scheme of the Azumaya $\mathbb{C}[X]$-algebra with involution (A, σ). Let

$\pi : \mathrm{Spin}_{A,\sigma} \to \mathrm{SO}_{A,\sigma}$ be the corresponding spinor X-group scheme. The X-group scheme $\mathrm{SO}_{A,\sigma}$ is the factor of $\mathrm{Spin}_{A,\sigma}$ modulo an inolution ε. Consider an involution τ on the variety $\mathrm{Spin}_{A,\sigma} \times \mathbb{C}^\times$, which takes any point (g, z) to the point $(\varepsilon(g), -z)$. Let Γ^+ be the factor variety of $\mathrm{Spin}_{A,\sigma} \times \mathbb{C}^\times$ modulo the involution τ. It is easy to check that Γ^+ is an X-group scheme. The map $\mathrm{Spin}_{A,\sigma} \times \mathbb{C}^\times \xrightarrow{pr_{\mathbb{C}^\times}} \mathbb{C}^\times \xrightarrow{\uparrow 2} \mathbb{C}^\times$ induces a map $Sn : \Gamma^+ \to \mathbb{C}^\times$ called the spinor norm map. For any section $\alpha : X \to \Gamma^+$ of the projection $\Gamma^+ \to X$ set $Sn(\alpha) = Sn \circ \alpha$. Let $a \in \mathbb{C}[X]^\times$ be a unite. Consider an equation

$$Sn(\alpha) = a. \tag{5}$$

If this equation has a solution over the field $\mathbb{C}(X)$, then it has a solution locally for the Zarisky topology on X. If $A = M_r(\mathbb{C}[X])$ and the involution σ corresponds to the quadratic space Q as above, then this result is equivalent to the result from the previous example.

Definition 3.1. Let X be as above. A smooth X-group scheme consists of the following data $p : \mathbf{G} \to X, \mu : \mathbf{G} \times_X \mathbf{G} \to \mathbf{G}, i : \mathbf{G} \to \mathbf{G}, e : X \to \mathbf{G}$, where p, μ, i, e are a regular maps. The requirements are these ones: p is smooth, μ is associative, e is a two-sided neutral "element" of the composition law μ, i is the inverse "element" for the composition law μ, e is a section of p.

If $\mathbf{G}_0 \subset \mathrm{GL}_n(\mathbb{C})$ is a linear complex algebraic group, then $\mathbf{G} = X \times \mathbf{G}_0$ with the obvious regular maps $p = pr_X, \mu, i, e$ form an X-group scheme. We say that an X-group scheme $\mathbf{G}$ is *holomorphically isomorphic* to the X-group scheme $X \times \mathbf{G}_0$ locally for the complex topology, if for any point $x \in X$ there is a holomorphic isomorphism $\varphi : \mathbf{G}|_U = p^{-1}(U) \to U \times \mathbf{G}_0$, which respects to the projection on U and to all the group data on both sides.

An X-group scheme $\mathbf{G}$ is called *a reductive (respectively, semi-simple; respectively, simple)* if it is an affine complex algebraic variety and for certain complex algebraic reductive group $\mathbf{G}_0$ it is holomorphically isomorphic to the X-group scheme $X \times \mathbf{G}_0$ locally for the complex topology on X. Recall that $\mathbf{G}_0$ is *required to be connected* .

This definition in the case of smooth affine complex algebraic variety X coincides with the one from Demazure and Grothendieck [1970b][Exp. XIX, Defn.2.7]. The class of reductive group schemes contains the class of semi-simple group schemes which in turn contains the class of simple group schemes. This notion of a simple X-group scheme coincides with the notion of a simple semi-simple X-group scheme from Demazure and Grothendieck [ibid.][Exp. XIX, Def. 2.7 and Exp. XXIV, 5.3].

Definition 3.2. Let $\mathbf{G}$ be a reductive X-group scheme. *A principal* $\mathbf{G}$*-bundle over* X consists of data $(p : E \to X, \nu : \mathbf{G} \times_X E \to E)$ such that p is a smooth surjective

regular map, ν is a $\mathbf{G}$-action on E respecting to the projections on X and
1) the regular map $\mathbf{G} \times_X E \to E \times_X E$ taking (g, e) to (ge, e) is an isomorphism of algebraic varieties;
2) for any point $x \in X$ there are a neighborhood V of x in the complex topology on X and an isomorphism of complex holomorphic varieties $\varphi : E|_V := p^{-1}(V) \to \mathbf{G}|_V$ such that φ respects to the projection on V and φ respects to the obvious left $\mathbf{G}$-actions on both sides.

A principal $\mathbf{G}$-bundle E *is called trivial* if there is an isomorphism $E \to \mathbf{G}$ over X, which respects to the obvious left $\mathbf{G}$-action on both sides. It is easy to check that E *is trivial* if and only if there is a section $s : X \to E$ of the projection $p : E \to X$.

In the Equation (1) in this section the X-group scheme is defined by the equation $T_1^2 - aT_2^2 = 1$. In the example (2) the X-group scheme is defined by the equation $N_{\mathbb{C}[X']/\mathbb{C}[X]}(\alpha) = 1$. In the example (3) the X-group scheme is defined by the equation $T_1^2 - aT_2^2 - bT_3^2 + abT_4^2 = 1$. In the example (4) the X-group scheme is defined by the equation $Nrd(\alpha) = 1$. In the example (5) the X-group scheme is $Spin_Q$. In the example (6) the X-group scheme is $Spin_{A,\sigma}$. The corresponding principal homogeneous bundles are described in examples (1)-(6). All these and many other examples illustrating the conjecture are simple consequences or partial cases of the following general result.

Theorem 3.3 (Fedorov and I. Panin [2015])**.** Let X be as above in this section. Let $\mathbf{G}$ be a reductive X-group scheme and E be a principal $\mathbf{G}$-bundle. Suppose for a non-zero function f the principal $\mathbf{G}$-bundle $E|_{X_f}$ is trivial over X_f. Then it is trivial locally for the Zarisky topology on X. That is for any point $x \in X$ there is a function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the principal $\mathbf{G}$-bundle $E|_{X_g}$ is trivial over X_g.

Remark 3.4. Let us point out that *the author still does not know* any proof of the result from the example (6), different from deriving it from Theorem 3.3. All results from examples (1)–(5) do have proofs avoiding any reference to Theorem 3.3.

Corollary 3.5 (of Theorem 3.3)**.** Let X be as above in this section. Let $\mathbf{G}$ be a reductive X-group scheme and E_1, E_2 be two principal $\mathbf{G}$-bundles. Suppose for a non-zero $f \in \mathbb{C}[X]$ the principal $\mathbf{G}$-bundles $E_1|_{X_f}$ and $E_2|_{X_f}$ over X_f. Then the principal $\mathbf{G}$-bundles E_1 and E_2 are isomorphic locally for the Zarisky topology on X.

Indeed, consider an X-group scheme $\underline{\mathrm{Aut}}_{\mathbf{G}}(E_1)$ of the $\mathbf{G}$-bundle automorphisms and an X-scheme of principal $\mathbf{G}$-bundle isomorphisms $\underline{\mathrm{Iso}}_{\mathbf{G}}(E_1, E_2)$. The latter scheme is a principal $\underline{\mathrm{Aut}}_{\mathbf{G}}(E_1)$-bundle and the X-group scheme $\underline{\mathrm{Aut}}_{\mathbf{G}}(E_1)$ is a reductive X-group scheme isomorphic to the X-group scheme $\mathbf{G}$ locally for the complex topology on X. A principal $\mathbf{G}$-bundle isomorphism $\varphi : E_1|_{X_f} \to E_2|_{X_f}$ is a section of $\underline{\mathrm{Iso}}_{\mathbf{G}}(E_1, E_2)$ over X_f. Hence $\underline{\mathrm{Iso}}_{\mathbf{G}}(E_1, E_2)$ has sections locally for the Zarisky topology on X. Thus the principal $\mathbf{G}$-bundles E_1 and E_2 are isomorphic locally for the Zarisky topology on X.

Corollary 3.6 (of Theorem 3.3)**.** Let $\mathbf{G}_1, \mathbf{G}_2$ be two simple X-group schemes of the type G_2 (resp. F_4, resp. E_8). That is the fibres and $\mathbf{G}_1, \mathbf{G}_2$ over a point $x \in X$ are of the type G_2 (resp. F_4, resp. E_8). Suppose for a non-zero element $f \in \mathbb{C}[X]$ the X_f-group schemes $(\mathbf{G_1})|_{X_f}$ and $(\mathbf{G_2})|_{X_f}$ are isomorphic. Then the X-group schemes $\mathbf{G_1}$ and $\mathbf{G_2}$ are isomorphic locally for the Zarisky topology on X.

Indeed, consider an X-scheme $\underline{\mathrm{Iso}}_{X-gr-sch}(\mathbf{G_1}, \mathbf{G_2})$. There is a regular map

$$\mathbf{G}_1 \times_X \underline{\mathrm{Iso}}_{X-gr-sch}(\mathbf{G_1}, \mathbf{G_2}) \to \underline{\mathrm{Iso}}_{X-gr-sch}(\mathbf{G_1}, \mathbf{G_2})$$

given by $(g_1, \varphi) \mapsto \varphi \circ \mathrm{conj}(g_1)$. This map makes the X-scheme $\underline{\mathrm{Iso}}_{X-gr-sch}(\mathbf{G_1}, \mathbf{G_2})$ a principal $\mathbf{G_1}$-bundle over X. This principal $\mathbf{G_1}$-bundle has a section over X_f. Hence for any point $x \in X$ there is a function $g \in \mathbb{C}[X]$ such that $g(x) \neq 0$ and the X-scheme $\underline{\mathrm{Iso}}_{X-gr-sch}(\mathbf{G_1}, \mathbf{G_1})$ has a section over X_g. The latter means that the X_g-group schemes $(\mathbf{G_1})|_{X_g}$ and $(\mathbf{G_2})|_{X_g}$ are isomorphic.

4 The conjecture, main results and some corollaries

Assume that U is a regular scheme. Let $\mathbf{G}$ be a reductive U-group scheme, that is, $\mathbf{G}$ is affine and smooth as a U-scheme and, moreover, the geometric fibers of $\mathbf{G}$ are connected reductive algebraic groups (see Demazure and Grothendieck [1970b][Exp. XIX, Definition 2.7]). Recall that a U-scheme E with an action of $\mathbf{G}$ is called a principal $\mathbf{G}$-bundle over U, if $\mathbf{G}$ is faithfully flat and quasi-compact over U and the action is simply transitive, that is, the natural morphism $\mathbf{G} \times_U E \to E \times_U E$ is an isomorphism. It is well known that such a bundle is trivial locally in the étale topology but in general not in the Zariski topology. Grothendieck and Serre conjectured that if E is generically trivial, then it is locally trivial in the Zariski topology (see Serre [1958][Remarque, p. 31], Grothendieck [1958][Remarque 3, pp. 26–-27], and Grothendieck [1968a][Remarque 1.11.a]). More precisely, the following conjecture is widely attributed to them.

Conjecture 1. Let R be a regular local ring, let K be its field of fractions. Let $\mathbf{G}$ be a reductive group scheme over $U := Spec R$, let E be a principal $\mathbf{G}$-bundle. If E is trivial over $Spec K$, then it is trivial. That is $E(R) \neq \varnothing$.

Theorem 4.1 (Main)**.** If R is a regular local ring containing a field, then the above conjecture holds.

This theorem is proved by R. Fedorov and the author in Fedorov and I. Panin [2015] in the case, when R contains an infinite field. It is proved by the author in I. Panin [2017b], when R contains a finite field.

Corollary 4.2 (of Theorem 4.1)**.** Let R be a regular local ring containing a field and $\mathbf{G}$ be a reductive R-group scheme. Let E_1, E_2 be two principal $\mathbf{G}$-bundles. Suppose they are isomorphic over the fraction field of R. Then they are isomorphic.

The proof literally repeats the proof of Corollary 3.5.

Corollary 4.3 (of Theorem 4.1)**.** Let R be a regular local ring containing a field and $\mathbf{G}$ be a reductive R-group scheme. Let $\mu : \mathbf{G} \to \mathbf{T}$ be a group scheme morphism to an R-torus $\mathbf{T}$ such that μ is locally in the étale topology on $Spec R$ surjective. Assume further that the R-group scheme $\mathbf{H} := \mathrm{Ker}(\mu)$ is reductive. Let K be the fraction field of R. Then the group homomorphism

$$\mathbf{T}(R)/\mu(\mathbf{G}(R)) \to \mathbf{T}(K)/\mu(\mathbf{G}(K)).$$

is injective.

To derive this corollary from Theorem 4.1 consider a commutative diagram

$$\begin{array}{ccccc} \mathbf{G}(R) & \xrightarrow{\mu} & \mathbf{T}(R) & \xrightarrow{\nu} & H^1_{et}(R, \mathbf{H}) \\ \downarrow & & \downarrow & & \downarrow \\ \mathbf{G}(K) & \xrightarrow{\mu} & \mathbf{T}(K) & \xrightarrow{\nu} & H^1_{et}(K, \mathbf{H}) \end{array}$$

By Theorem 4.1 the right vertical arrow has trivial kernel. Now a simple diagram chase completes the proof. The latter corollary extends all the known results of this form proved in Colliot-Thélène and Ojanguren [1992], I. A. Panin and Suslin [1997], Zaĭnullin [2000], and Ojanguren, I. Panin, and Zainoulline [2004].

Corollary 4.4 (of Theorem 4.1)**.** Under the notation and the hypothesis of the previous corollary the following sequence is exact

$$\{1\} \to \mathbf{T}(R)/\mu(\mathbf{G}(R)) \to \mathbf{T}(K)/\mu(\mathbf{G}(K)) \xrightarrow{\Sigma\, res_p} \oplus_{ht(p)=1} \mathbf{T}(K)/\mathbf{T}(R_p)\cdot\mu(\mathbf{G}(K)) \to \{1\},$$

where p runs all hight 1 prime ideals in R and res_p is the obvious map.

The exactness at the term $\mathbf{T}(R)/\mu(\mathbf{G}(R))$ is due to the previous corollary. The surjectivity of the map $\Sigma\, res_p$ is due to Colliot-Thélène and Sansuc [1987]. The exactness at the middle term is proved in I. A. Panin [2016a], I. Panin [2017c].

There are two other very general results concerning the conjecture: due to Nisnevič [1977] and due to Colliot-Thélène and Sansuc [1987] (see the history of the topic).

5 History of the topic

History of the topic. — In his 1958 paper Jean-Pierre Serre asked whether a principal bundle is Zariski locally trivial, once it has a rational section (see Serre [1958][Remarque, p. 31]). In his setup the group is any algebraic group over an algebraically closed field. He gave an affirmative answer to the question when the group is PGL(n) (see Serre [ibid.][Prop. 18]) and when the group is an abelian variety (see Serre [ibid.][Lemme 4]). In the same year, Alexander Grothendieck asked a similar question (see Grothendieck [1958][Remarque 3, pp. 26–27]). A few years later, Grothendieck conjectured that the statement is true for any semi-simple group scheme over any regular local scheme (see Grothendieck [1965][Remarque 1.11.a]). Now by the Grothendieck–Serre conjecture we mean Conjecture 1 though this may be slightly inaccurate from historical perspective. Many results corroborating the conjecture are known.

Here is a list of known results in the same vein, corroborating the Grothendieck–Serre conjecture.

• The case when the group is PGL_n and the base field is algebraically closed is done by J.-P. Serre in 1958 in Serre [1958, Prop. 18].

• The case when the group scheme is PGL_n and the ring R is an arbitrary regular local ring is done by A. Grothendieck in 1968 in Grothendieck [1968a].

• The case when the local ring R contains a field of characteristic not 2 the group is SO_n over the ground field is done by M. Ojanguren in 1982 in Ojanguren [1980].

• The case of an arbitrary reductive group scheme over a discrete valuation ring or over a henselian ring is solved by Nisnevič [1977] in 1984.

• The case, where $\mathbf{G}$ is an arbitrary torus over a regular local ring, was settled by Colliot-Thélène and Sansuc [1987] in 1987.

• The case, when $\mathbf{G}$ is quasi-split reductive group scheme over arbitrary two-dimensional local rings, is solved by Nisnevich [1984] in 1989.

• The case, where the group scheme $\mathbf{G}$ comes from an infinite perfect ground field, solved by Colliot-Thélène and Ojanguren [1992] in 1992. As far as we know this work was inspired by the one Ojanguren [1980].

• The case, where the group scheme $\mathbf{G}$ comes from an arbitrary infinite ground field, solved by Raghunathan [1994, 1995] in 1994;

• O. Gabber announced in 1994 a proof for group schemes coming from arbitrary ground fields (including finite fields).

• For the group scheme $\mathrm{SL}_{1,A}$, where A is an Azumaya R-algebra and R contains a field the conjecture is solved by I. A. Panin and Suslin [1997] in 1998.

• For the unitary group scheme $\mathrm{U}^{\varepsilon}_{A,\sigma}$, where (A,σ) is an Azumaya R-algebra with involution R contains a field of characteristic not 2 the conjecture is solved by Ojanguren and I. Panin [2001] in 2001.

• For the special unitary group scheme $\mathrm{SU}_{A,\sigma}$, where (A,σ) is an Azumaya R-algebra with a unitary involution and R contains a field of characteristic not 2 the conjecture is solved by Zaĭnullin [2000] in 2001.

• For the spinor group scheme Spin_Q of a quadratic space Q over R containing a field of characteristic not 2 the conjecture is solved Ojanguren, I. Panin, and Zainoulline [2004] in 2004.

• Under an isotropy condition on $\mathbf{G}$ the conjecture is proved by A. Stavrova, N. Vavilov and the author in a series of preprints in 2009, published as papers in 2015 in I. Panin, A. Stavrova, and Vavilov [2015a] and in 2016 in I. A. Panin [2016a].

• The case of strongly inner simple adjoint group schemes of the types E_6 and E_7 is done by the second author, V. Petrov, A. Stavrova and the second author in 2009 in I. Panin, Petrov, and A. Stavrova [2009]. No isotropy condition is imposed there.

• The case, when $\mathbf{G}$ is of the type F_4 with trivial f_3-invariant and the field is infinite and perfect, is settled by Petrov and A. Stavrova [2009] in 2009.

• The case, when $\mathbf{G}$ is of the type F_4 with trivial g_3-invariant and the field is of characteristic zero, is settled by Chernousov [2010] in 2010;

• The conjecture is solved when R contains an infinite field, by R. Fedorov and the author in a preprint in 2013 and published in Fedorov and I. Panin [2015] in 2015.

• The conjecture is solved by the author in the case, when R contains a finite field in I. Panin [2015] (for a better structured proof see I. Panin [2017b]).

So, *the conjecture is solved in the case, when R contains a field*.

The case of mixed characteristic is widely open. Let us indicate two recent interesing preprints Fedorov [2015] and I. A. Panin and A. K. Stavrova [2016]. In Fedorov [2015] the conjecture is solved for a large class of regular local rings of mixed characteristic assuming that $\mathbf{G}$ splits. In I. A. Panin and A. K. Stavrova [2016] the conjecture is solved for any semi-local Dedekind domain providing that $\mathbf{G}$ is simple simply-connected and $\mathbf{G}$ contains a torus $\mathbb{G}_{m,R}$.

6 Sketch of the proof of Theorem 4.1 for an infinite field

To escape technicalities we suppose also that the field k below *is algebraically closed* (for instance, k is the complex numbers $\mathbb{C}$). And also suppose that the ring R is the semi-local ring $\mathcal{O}_{X,x_1,\dots,x_n}$ of finitely many closed points on a smooth affine k-variety X. There are two very general purity theorems I. A. Panin [2016a, Thm. 1.0.1, Thm. 1.0.2] which allows I. A. Panin [ibid., Thm. 1.0.3] to reduce Theorem 4.1 to the case, when the group scheme $\mathbf{G}$ is semi-simple simply-connected. Then using standard arguments as in I. Panin, A. Stavrova, and Vavilov [2015b] one can reduce the latter case to the case of simple and simply-connected group scheme $\mathbf{G}$ (*point out that this reduction requires to work with semi-local rings*). So, we will consider below only the case of simple and simply-connected group scheme **G**. And for simplicity of notation we will suppose below the ring R is the local ring $\mathcal{O}_{X,x}$ of a closed point x on a smooth affine k-variety X.

Theorem 6.1 (I. Panin, A. Stavrova, and Vavilov [ibid.]). Let R be the local ring of a closed point on an irreducible smooth affine variety over the field k, set $U = \operatorname{Spec} R$. Let **G** be a simple simply-connected group scheme over U (see Demazure and Grothendieck [1970a, Exp. XXIV, Sect. 5.3] for the definition). Let $\mathcal{G}$ be a principal **G**-bundle over U which is trivial over the principal open subset $U_f \subset U$ for a non-zero $f \in R$. Then there exists a principal **G**-bundle E_t over $\mathbb{A}^1_U$ and a monic polynomial $h(t) \in R[t]$ such that
(i) the **G**-bundle E_t is trivial over $(\mathbb{A}^1_U)_h$,
(ii) the evaluation of E_t at $t = 0$ coincides with the original **G**-bundle $\mathcal{G}$,
(iii) $\{1\} \times U \subset (\mathbb{A}^1_U)_h$.

Remark 6.2. If the field $k = \mathbb{C}$, then the principal **G**-bundle E_t regarded as a topological principal **G**-bundle for the complex topology is of the form $p^*(E_0)$ for a topological principal **G**-bundle E_0, where $p : \mathbb{A}^1_U \to U$ is the projection. By the item (iii) and (ii) of the latter theorem the principal **G**-bundle $\mathcal{G}$ *regarded as a topological principal* **G***-bundle is trivial.* Hence it is trivial even as the complex holomorphic principal **G**-bundle.

However *these kind of arguments do not work in general for algebraic principal* **G***-bundles* since there are principal **G**-bundles on $\mathbb{A}^1_U$ which do not come from U (see Fedorov [2016]). If $k = \mathbb{R}$ then there are many examples of principal **G**-bundles on $\mathbb{A}^1_U$ which do not come from U. Those examples can be deduced from Knus, R. Parimala, and Sridharan [1981/82], Ojanguren, R. Parimala, and Sridharan [1983], S. Parimala [1978].

That is why we need in the following proposition and theorem.

Proposition 6.3 (Fedorov and I. Panin [2015]). Let k,R, U, **G**, be the same as in Theorem 6.1. Let $Z \subset \mathbb{A}^1_U$ be a closed subscheme finite over U. Then there exists a closed subscheme Y in $\mathbb{A}^1_U$ such that Y is étale and finite over U, the Y-group scheme $\mathbf{G}_Y := \mathbf{G}\times_U Y$ is quasi-split and $Y \cap Z = \varnothing$.

Theorem 6.4 (Fedorov and I. Panin [ibid.]). Let k, R, U, **G**, be the same as in Theorem 6.1. Let $Z \subset \mathbb{A}^1_U$ be a closed subscheme finite over U. Let $Y \subset \mathbb{A}^1_U$ be a closed subscheme étale and finite over U. Assume that $Y \cap Z = \varnothing$, and $\mathbf{G}_Y := \mathbf{G} \times_U Y$ is quasi-split.

Let E be a principal **G**-bundle over $\mathbb{P}^1_U$ such that its restriction to $\mathbb{P}^1_U - Z$ is trivial. Then the restriction of E to $\mathbb{P}^1_U - Y$ is also trivial.

Derive now the simple simply-connected case (geometric) of Theorem 4.1 from these three statements. Let $\mathcal{G}$ be a principal **G**-bundle over U which is trivial over the principal open subset $U_f \subset U$ for a non-zero $f \in R$. By Theorem 6.1 there are a monic polinomial $h(t) \in R[t]$ and a principal **G**-bundle E_t over $\mathbb{A}^1_U$ such that (i) and (ii) hold. Since E_t is trivial over $(\mathbb{A}^1_U)_{h(t)}$ there is a principal **G**-bundle E over $\mathbb{P}^1_U$ such that $E|_{\mathbb{P}^1_U - Z}$ is trivial and $E_{\mathbb{A}^1_U} = E_t$ (here $Z \subset \mathbb{A}^1_U \subset \mathbb{P}^1_U$ is the vanishing locus of $h(t) = 0$). By Proposition 6.3 there is a closed subscheme Y in $\mathbb{A}^1_U$ such that Y is étale and finite over U, the Y-group scheme $\mathbf{G}_Y := \mathbf{G} \times_U Y$ is quasi-split and $Y \cap Z = \varnothing$. By Theorem 6.4 the restriction of E to $\mathbb{P}^1_U - Y$ is trivial. Since U is local for each section $s : U \to \mathbb{A}^1_U$ of the projection $\mathbb{A}^1_U \to U$ either $s(U) \cap Y = \varnothing$ or $s(U) \cap Z = \varnothing$. In any case, the principal $\mathbf{G}$-bundle $s^*(E_t) = s^*(E)$ is trivial. By the item (ii) of Theorem 6.1 the original $\mathbf{G}$-bundle $\mathcal{G}$ is trivial.

First, we give a sketch of the proof of Theorem 6.1. Let k be the field. Let X be an affine k-smooth irreducible k-variety, and let x be a closed point in X. Let $U = Spec(\mathcal{O}_{X,x})$ and f $\in k[X]$ be a non-zero function vanishing the point x. Let $\mathbf{G}_X$ be a simple simply-connected group scheme over X, $\mathbf{G}$ be its restriction to U.

Beginning with these data it is constructed in I. Panin, A. Stavrova, and Vavilov [2015b] a monic polinomial $h \in \mathcal{O}_{X,x}[t]$, a commutative diagram of schemes with the irreducible affine U-smooth variety Y

$$\begin{array}{ccccc} (\mathbb{A}^1 \times U)_h & \xleftarrow{\ \tau_h\ } & Y_h := Y_{\tau^*(h)} & \xrightarrow{(p_X)|_{Y_h}} & X_f \\ \big\downarrow{\scriptstyle inc} & & \big\downarrow{\scriptstyle inc} & & \big\downarrow{\scriptstyle inc} \\ (\mathbb{A}^1 \times U) & \xleftarrow{\ \tau\ } & Y & \xrightarrow{\ p_X\ } & X \end{array} \tag{6}$$

and a morphism $\delta : U \to Y$ subjecting to the following conditions:

(a) the left hand side square is *an elementary distinguished square* in the category of affine U-smooth schemes in the sense of Morel and Voevodsky [1999, Defn.3.1.3];

(b) $p_X \circ \delta = can : U \to X$, where can is the canonical morphism;

(c) $\tau \circ \delta = i_0 : U \to \mathbb{A}^1 \times U$ is the zero section of the projection $pr_U : \mathbb{A}^1 \times U \to U$;

(d) $h(1) \in \mathcal{O}_{X,x}[t]$ is a unit;

(e) for $p_U := pr_U \circ \tau$ there is a Y-group scheme isomorphism $\Phi : p_U^*(\mathbf{G}) \to p_X^*(\mathbf{G}_X)$ with $\delta^*(\Phi) = id_{\mathbf{G}}$.

Given this geometric result a proof of Theorem 6.1 run as follows. In general, $\mathbf{G}$ does not come from X. However we may assume, that $\mathbf{G}$ is a restriction to U of a simple and simply-connected X-group scheme $\mathbf{G}_X$, $\mathcal{G}$ is defined over X. Say, let $\mathcal{G}'$ be a principal $\mathbf{G}_X$-bundle on X with $\mathcal{G}'|_U = \mathcal{G}$ and such that $\mathcal{G}'$ is trivial over X_f for an $0 \neq f \in k[X]$ with $f(x) = 0$. In this case there are two reductive group schemes on Y. Namelly, $p_U^*(\mathbf{G})$ and $p_X^*(\mathbf{G}_X)$. By the property (b) they coincides when restricted to $\delta(U)$. By the property (e) the scheme Y is chosen such that two reductive group schemes $p_U^*(\mathbf{G})$ and $p_X^*(\mathbf{G}_X)$

on Y are isomorphic via an Y-group scheme isomorphism Φ and the restriction of Φ to $\delta(U)$ is the identity. Take $p_X^*(\mathcal{G}')$ and regard it as a principal $p_U^*(\mathbf{G})$-bundle using the isomorphism Φ. Denote that principal $p_U^*(\mathbf{G})$-bundle ${}_U p_X^*(\mathcal{G})$. It is trivial on Y_h, since $p_X^*(\mathcal{G}')$ is trivial on Y_h. Take the trivial $pr_U^*(\mathbf{G})$-bundle on $\mathbb{A}^1 \times U$ and glue it with ${}_U p_X^*(\mathcal{G})$ via an isomorphism over Y_h. That is possible by the condition (a). This way we get a principal $\mathbf{G}$-bundle $\mathcal{G}_t$ over $\mathbb{A}^1 \times U$ which is trivial over the open subscheme $(\mathbb{A}^1 \times U)_h$ and such that ${}_U p_X^*(\mathcal{G}) = \tau^*(\mathcal{G}_t)$. Clearly, it is the desired one. The polinomial h is the polinomial above. Whence the Theorem 6.1.

Secondly, we give a sketch of the proof of Proposition 6.3. Let $\mathfrak{B}$ be the U-scheme of Borel subgroup schemes of $\mathbf{G}$. It is a smooth projective U-scheme (see Demazure and Grothendieck [1970a, Cor. 3.5, Exp. XXVI]). Take an embedding of $\mathfrak{B}$ into a projective space $\mathbb{P}_U^N$ and intersect $\mathfrak{B}$ with appropriately chosen family of hyperplanes. Arguing as in the proof of Ojanguren and I. Panin [2001, Lemma 7.2], we get a scheme Y finite and étale over U and such that the Y-group scheme $\mathbf{G}_Y := \mathbf{G} \times_U Y$ is quasi-split. Since the field k is infinite and Y is finite étale over U, we can choose a closed U-embedding of Y in $\mathbb{A}_U^1$. We will identify Y with the image of this closed embedding. Since Y is finite over U, it is closed in $\mathbb{P}_U^1$. Applying to Y an appropriative affine U-transformation of $\mathbb{A}_U^1$ we get Y such that $Y \cap Z = \varnothing$. Whence the Proposition 6.3

The next result is a partial case of Theorem 9.6 of I. Panin, A. Stavrova, and Vavilov [2015b].

Proposition 6.5. Let U be as above and let $u \in U$ be its closed point. Let $\mathcal{E}$ be a $\mathbf{G}$-bundle over $\mathbb{P}_U^1$ such that $\mathcal{E}|_{\mathbb{P}_u^1}$ is a trivial $\mathbf{G}_u$-bundle. Assume that there exists a closed subscheme T of $\mathbb{P}_U^1$ finite over U such that the restriction of $\mathcal{E}$ to $\mathbb{P}_U^1 - T$ is trivial. Then E is of the form: $\mathcal{E} = \mathrm{pr}^*(\mathcal{E}_0)$, where $\mathcal{E}_0$ is a principal $\mathbf{G}$-bundle over U and $\mathrm{pr} : \mathbb{P}_U^1 \to U$ is the canonical projection.

If furthermore $T \cap \{\infty\} \times U = \varnothing$, then $\mathcal{E}$ is trivial.

Finally, we give a sketch of the proof of Theorem 6.4. Let $(Y^h, \pi : Y^h \to \mathbb{A}_U^1, s : Y \to Y^h)$ be the henselization of the pair $(\mathbb{A}_U^1, Y)$. Here $s : Y \to Y^h$ is the canonical closed embedding, see Fedorov and I. Panin [2015, Sect. 5.3] for more details. Let $in : \mathbb{A}_U^1 \to \mathbb{P}_U^1$ be the embedding. Set $\dot{Y}^h := Y^h - s(Y)$. Note that as Y^h, so $\dot{Y}^h$ *are affine schemes*, see Fedorov and I. Panin [ibid., Sect. 5.3, Prop. 5.13]. Consider the following cartesian square of schemes

$$\begin{array}{ccc} \dot{Y}^h & \xrightarrow{j} & Y^h \\ \downarrow & & \downarrow{\scriptstyle in\circ\pi} \\ \mathbb{P}_U^1 - Y & \xrightarrow{i} & \mathbb{P}_U^1 \end{array} \tag{7}$$

As explained in Fedorov and I. Panin [2015, Prop. 5.15, Constr. 5.16] that square can be used to get $\mathbf{G}$-bundles on $\mathbb{P}^1_U$ beginning with a $\mathbf{G}$-bundle on $\mathbb{P}^1_U - Y$ and its trivialization over $\dot{Y}^h$. Let E' be a $\mathbf{G}$-bundle over $\mathbb{P}^1_U - Y$. Denote by $\mathrm{Gl}(E', \varphi)$ the $\mathbf{G}$-bundle over $\mathbb{P}^1_U$ obtained by gluing E' with the trivial $\mathbf{G}$-bundle $\mathbf{G} \times_U Y^h$ via a $\mathbf{G}$-bundle isomorphism $\varphi : \mathbf{G} \times_U \dot{Y}^h \to E'|_{\dot{Y}^h}$.

Similarly, set $Y_u := Y \times_U u$ and denote by Y^h_u the henselization of the pair $(\mathbb{A}^1_u, Y_u)$, let $s_u : Y_u \to Y^h_u$ be the closed embedding. Set $\dot{Y}^h_u := Y^h_u - s_u(Y_u)$. Let E'_u be a $\mathbf{G}_u$-bundle over $\mathbb{P}^1_u - Y_u$, where $\mathbf{G}_u := \mathbf{G} \times_U u$. Denote by $\mathrm{Gl}_u(E'_u, \varphi_u)$ the $\mathbf{G}_u$-bundle over $\mathbb{P}^1_u$ obtained by gluing E'_u with the trivial bundle $\mathbf{G}_u \times_u Y^h_u$ via a $\mathbf{G}_u$-bundle isomorphism $\varphi_u : \mathbf{G}_u \times_u \dot{Y}^h_u \to E'_u|_{\dot{Y}^h_u}$.

Note that the $\mathbf{G}$-bundle E can be presented in the form $\mathrm{Gl}(E', \varphi)$, where $E' = E|_{\mathbb{P}^1_U - Y}$. The idea is to show that

(∗) *There is $\alpha \in \mathbf{G}(\dot{Y}^h)$ such that the $\mathbf{G}_u$-bundle $Gl(E', \varphi \circ \alpha)|_{\mathbb{P}^1_u}$ is trivial (here α is regarded as an automorphism of the $\mathbf{G}$-bundle $\mathbf{G} \times_U \dot{Y}^h$ given by the right translation by the element α).*

If we find α satisfying condition (∗), then Proposition 6.5, applied to $T = Y \cup Z$, shows that the $\mathbf{G}$-bundle $\mathrm{Gl}(E', \varphi \circ \alpha)$ is trivial over $\mathbb{P}^1_U$. On the other hand, its restriction to $\mathbb{P}^1_U - Y$ coincides with the $\mathbf{G}$-bundle $E' = E|_{\mathbb{P}^1_U - Y}$. *Thus $E|_{\mathbb{P}^1_U - Y}$ is a trivial $\mathbf{G}$-bundle*
.

To prove (∗), one should show that

(i) the bundle $E|_{\mathbb{P}^1_u - Y_u}$ is trivial;

(ii) each element $\gamma_u \in \mathbf{G}_u(\dot{Y}^h_u)$ can be written in the form $\gamma_u = \alpha|_{\dot{Y}^h_u}$ for a certain element $\alpha \in \mathbf{G}(\dot{Y}^h)$.

If we succeed to show (i) and (ii), then we proceed as follows. Present the $\mathbf{G}$-bundle E in the form $\mathrm{Gl}(E', \varphi)$ as above. Observe that

$$\mathrm{Gl}(E', \varphi)|_{\mathbb{P}^1_u} \cong \mathrm{Gl}_u(E'_u, \varphi_u),$$

where $E'_u := E'|_{\mathbb{P}^1_u - Y_u}$, $\varphi_u := \varphi|_{\mathbf{G}_u \times_u \dot{Y}^h_u}$.

Using property (i), find an element $\gamma_u \in \mathbf{G}_u(\dot{Y}^h_u)$ such that the $\mathbf{G}_u$-bundle $\mathrm{Gl}_u(E'_u, \varphi_u \circ \gamma_u)$ is trivial. For this γ_u find an element α as in (ii). Finally take the $\mathbf{G}$-bundle $\mathrm{Gl}(E', \varphi \circ \alpha)$. Then its restriction to $\mathbb{P}^1_u$ is trivial. Indeed, one has a chain of $\mathbf{G}_u$-bundle isomorphisms

$$\mathrm{Gl}(E', \varphi \circ \alpha)|_{\mathbb{P}^1_u} \cong \mathrm{Gl}_u(E'_u, \varphi_u \circ \alpha|_{\dot{Y}^h_u}) = \mathrm{Gl}_u(E'_u, \varphi_u \circ \gamma_u),$$

which is trivial by the very choice of γ_u. Thus, (∗) will be achieved.

Let us prove (i) and (ii). The $\mathbf{G}_u$-bundle E_u is trivial over $\mathbb{P}^1_u - Z_u$. The field $k(u) = k$ is algebraically closed. By a theorem Grothendieck–Harder Harder [1968, Satz 3.1, 3.4] there is an algebraic group morphism $\lambda : \mathbb{G}_{m,k(u)} \to \mathbf{G}_u$ such that the $\mathbf{G}_u$-bundle E_u on $\mathbb{P}^1_u$ is isomorphic to the one $\mathbf{G}_u \times_{\mathbb{G}_{m,k(u)}} \mathcal{O}(-1)$, where $\mathcal{O}(-1)$ is the $\mathbb{G}_{m,k(u)}$-bundle

$\mathbb{A}^2_{k(u)} - 0 \to \mathbb{P}^1_{k(u)}$. Since the field k is algebraically closed, hence $\mathbb{P}^1_u - Y_u$ is contained in an affine line $\mathbb{A}^1_u$. Thus the restrictions of $\mathcal{O}(-1)$ and E_u to $\mathbb{P}^1_u - Y_u$ are trivial. So, (i) is achieved.

To complete the proof it remains to achieve (ii). By our assumption on Y, the group scheme $\mathbf{G}_Y = \mathbf{G} \times_U Y$ is quasi-split. Thus we can and will choose a Borel subgroup scheme $\mathbf{B}^+$ in $\mathbf{G}_Y$.

Since Y is an affine scheme, by Demazure and Grothendieck [1970a, Exp. XXVI, Cor. 2.3, Th 4.3.2(a)] there is an opposite to $\mathbf{B}^+$ Borel subgroup scheme $\mathbf{B}^-$ in $\mathbf{G}_Y$. Let $\mathbf{U}^+$ be the unipotent radical of $\mathbf{B}^+$, and let $\mathbf{U}^-$ be the unipotent radical of $\mathbf{B}^-$.

We will write $\mathbf{E}$ for the functor, sending a Y-scheme T to the subgroup $\mathbf{E}(T)$ of the group $\mathbf{G}_Y(T) = \mathbf{G}(T)$ generated by the subgroups $\mathbf{U}^+(T)$ and $\mathbf{U}^-(T)$ of the group $\mathbf{G}_Y(T) = \mathbf{G}(T)$.

The functor $\mathbf{E}$ has the property that for every closed subscheme S in an affine Y-scheme T *the induced map* $\mathbf{E}(T) \to \mathbf{E}(S)$ *is surjective.* Indeed, the restriction maps $\mathbf{U}^\pm(T) \to \mathbf{U}^\pm(S)$ are surjective, since $\mathbf{U}^\pm$ are isomorphic to vector bundles as Y-schemes (see Demazure and Grothendieck [ibid., Exp. XXVI, Cor. 2.5]).

Recall that (Y^h, π, s) is the henselization of the pair $(\mathbb{A}^1_U, Y)$. Also, $in : \mathbb{A}^1_U \to \mathbb{P}^1_U$ is the embedding. Denote the projection $\mathbb{A}^1_U \to U$ by pr and the projection $\mathbb{A}^1_Y \to Y$ by pr_Y. It is proved in Fedorov and I. Panin [2015, Lemma 5.25] the following

Claim. There is a morphism $r : Y^h \to Y$ making the following diagram commutative

$$
\begin{array}{ccc}
Y^h & \xrightarrow{r} & Y \\
{\scriptstyle in\circ\pi}\downarrow & & \downarrow{\scriptstyle pr|_Y} \\
\mathbb{P}^1_U & \xrightarrow{pr} & U
\end{array}
\tag{8}
$$

and such that $r \circ s = \mathrm{Id}_Y$.

We view Y^h as a Y-scheme via r. Thus various subschemes of Y^h also become Y-schemes. In particular, $\dot{Y}^h$ and $\dot{Y}^h_u$ are Y-schemes, and we can consider

$$\mathbf{E}(\dot{Y}^h) \subset \mathbf{G}(\dot{Y}^h) \text{ and } \mathbf{E}(\dot{Y}^h_u) \subset \mathbf{G}(\dot{Y}^h_u) = \mathbf{G}_u(\dot{Y}^h_u).$$

Since $\dot{Y}^h_u$ is an affine scheme corresponding to the direct product of few fields, $\mathbf{G}_u$ is *simply-connected and quasi-split*, hence one has an equality $\mathbf{G}_u(\dot{Y}^h_u) = \mathbf{E}(\dot{Y}^h_u)$. As indicated right above the diagram (7) the scheme $\dot{Y}^h$ is affine. Since $\dot{Y}^h_u$ is its closed subscheme the group homomorphism $\mathbf{E}(\dot{Y}^h) \to \mathbf{E}(\dot{Y}^h_u)$ is surjective. Thus the the group homomorphism $\mathbf{G}(\dot{Y}^h) \to \mathbf{G}(\dot{Y}^h_u) = \mathbf{G}_u(\dot{Y}^h_u)$ is surjective as well. We achieved the property (ii). Whence the Theorem 6.4. The sketch of the proof of the Theorem 4.1 is completed.

7 Sketch of the proof of Theorem 4.1 for a finite field

Let k be a finite field. Give a sketch of the proof of Theorem 4.1 in this case. The outline of the proof is the same. So, we will focus on crucial differences. There is a reduction to the case of simple and simply-connected group scheme $\mathbf{G}$ over the semi-local ring of finitely many closed points on a smooth affine variety X So, we will consider below only the case of simple and simply-connected group scheme $\mathbf{G}$. And for simplicity of notation we will suppose below the ring R is the local ring $\mathcal{O}_{X,x}$ of a closed point x on a smooth affine k-variety X.

The statement of Theorem 6.1 remains the same in the case of finite base field.

Theorem 7.1 (I. Panin [2017a])**.** Let R be the local ring of a closed point on an irreducible smooth affine variety over the finite field k, set $U = \text{Spec } R$. Let $\mathbf{G}$ be a simple simply-connected group scheme over U. Let $\mathcal{G}$ be a principal $\mathbf{G}$-bundle over U which is trivial over the principal open subset $U_f \subset U$ for a non-zero $f \in R$. Then there exists a principal $\mathbf{G}$-bundle E_t over $\mathbb{A}^1_U$ and a monic polynomial $h(t) \in R[t]$ such that

(i) the $\mathbf{G}$-bundle E_t is trivial over $(\mathbb{A}^1_U)_h$,

(ii) the evaluation of E_t at $t = 0$ coincides with the original $\mathbf{G}$-bundle $\mathcal{G}$,

(iii) $h(1) \in R$ is invertible in R.

Proposition 6.3 one needs to replace with the following one

Proposition 7.2 (I. Panin [2017b])**.** Let k,R, U, $\mathbf{G}$, be the same as in Theorem 6.1 and k be the finite field. Let $Z \subset \mathbb{A}^1_U$ be a closed subscheme finite over U. Then there exists a closed subscheme Y in $\mathbb{A}^1_U$ such that Y is étale and finite over U,

(a) the Y-group scheme $\mathbf{G}_Y := \mathbf{G} \times_U Y$ is quasi-split,

(b) $\text{Pic}(\mathbb{P}^1_u - Y_u) = 0$,

(c) $Y \cap Z = \varnothing$.

Theorem 6.4 one needs to replace with the following one

Theorem 7.3 (I. Panin [ibid.])**.** Let k, R, U, $\mathbf{G}$, be the same as in Theorem 6.1. Let $Z \subset \mathbb{A}^1_U$ be a closed subscheme finite over U. Let $Y \subset \mathbb{A}^1_U$ be a closed subscheme étale and finite over U. Assume that $Y \cap Z = \varnothing$, $\mathbf{G}_Y := \mathbf{G} \times_U Y$ is quasi-split and $\text{Pic}(\mathbb{P}^1_u - Y_u) = 0$.

Let E be a principal $\mathbf{G}$-bundle over $\mathbb{P}^1_U$ such that its restriction to $\mathbb{P}^1_U - Z$ is trivial. Then the restriction of E to $\mathbb{P}^1_U - Y$ is also trivial.

The derivation of the simple simply-connected case (geometric) of Theorem 4.1 from these three statements remains the same as in the infinite field case above.

The proof of Theorem 7.1 is essentially more involved than the proof of Theorem 6.1 and is based on some new ideas (see I. Panin [2017a]).

Give a sketch of the proof of Proposition 7.2. For the closed point $u \in U$ choose a Borel subgroup $\mathbf{B}_u$ in $\mathbf{G}_u$. The latter is possible since the field $k(u)$ is finite. Let $\mathcal{B}$ be the U-scheme of Borel subgroup schemes of $\mathbf{G}$. It is a smooth projective U-scheme (see Demazure and Grothendieck [1970a, Cor. 3.5, Exp. XXVI]). The subgroup $\mathbf{B}_u$ in $\mathbf{G}_u$ is a $k(u)$-rational point b in the fibre of $\mathcal{B}$ over the point u.

Applying several many times Poonen's Bertini type theorem Poonen [2004, Thm. 1.2] find a closed subscheme Y' of $\mathcal{B}$ such that Y' is étale over U and the point b is in Y'. Clearly, the Y-group scheme $\mathbf{G}_Y := \mathbf{G} \times_U Y$ is quasi-split. To finish the proof of Proposition 7.2 it remains to find a closed embedding of Y' into $\mathbb{A}^1_U$ which satisfies properties (b) and (c).

However it might happen that there is no closed embedding of Y' into $\mathbb{A}^1_U$ at all. Indeed, if the number of $k(u)$-rational point on Y'_u *is strictly more* than the cardinality of the field $k(u)$, than there is no closed embedding of Y' into $\mathbb{A}^1_U$ at all. *To avoid this trouble we need in the following*

Lemma 7.4 (I. Panin [2017b]). Let U be as in the Proposition 7.2. Let $Z \subset \mathbb{A}^1_U$ be a closed subscheme finite over U. Let $Y' \to U$ be a finite étale morphism such that for the closed point u in U the fibre Y'_u of Y' over u contains a $k(u)$-rational point. Then there are finite field extensions k_1 and k_2 of the finite field k such that
(i) the degrees $[k_1 : k]$ and $[k_2 : k]$ are coprime,
(ii) $k(u) \otimes_k k_r$ is a field for $r = 1$ and $r = 2$,
(iii) the degrees $[k_1 : k]$ and $[k_2 : k]$ are strictly greater than any of the degrees $[k(z) : k(u)]$, where z runs over all closed points of Z,
(iv) there is a closed embedding of U-schemes $Y'' = ((Y' \otimes_k k_1) \sqcup (Y' \otimes_k k_2)) \xrightarrow{i} \mathbb{A}^1_U$,
(v) for $Y = i(Y'')$ one has $Y \cap Z = \varnothing$,
(vi) for the closed point u in U one has $\mathrm{Pic}(\mathbb{P}^1_u - Y_u) = 0$.

Finish now the proof of the proposition. The U-scheme Y' satisfies the hypotheses of Lemma 7.4. Take the closed subscheme Y of $\mathbb{A}^1_U$ as in the item (v) of the Lemma. For that specific Y the conditions (b) and (c) of the Proposition are obviously satisfied. The condition (a) is satisfied too, since it is satisfied already for the U-scheme Y'. The proposition follows.

Finally, we give a sketch of the proof of Theorem 7.3. It almost literally repeats the sketch of the proof of Theorem 6.4. The only difference is in checking the triviality of the bundle $E|_{\mathbb{P}^1_u - Y_u}$.

The $\mathbf{G}_u$-bundle E_u is trivial over $\mathbb{P}^1_u - Z_u$. The field $k(u)$ is finite and the $k(u)$-group $\mathbf{G}_u$ is quasi-split. By a theorem due to Harder Harder [1968, Satz 3.1, 3.4] there is an algebraic group morphism $\lambda : \mathbb{G}_{m,k(u)} \to \mathbf{G}_u$ such that the $\mathbf{G}_u$-bundle E_u on $\mathbb{P}^1$ is isomorphic to the one $\mathbf{G}_u \times_{\mathbb{G}_{m,k(u)}} \mathcal{O}(-1)$, where $\mathcal{O}(-1)$ is the $\mathbb{G}_{m,k(u)}$-bundle $\mathbb{A}^2_{k(u)} - 0 \to \mathbb{P}^1_{k(u)}$. Since $\mathrm{Pic}(\mathbb{P}^1_u - Y_u) = 0$, hence the restriction of $\mathcal{O}(-1)$ to $\mathbb{P}^1_u - Y_u$

is trivial. Thus, so is the restriction of E_u to $\mathbb{P}^1_u - Y_u$. Theorem 4.1 for the finite field case is proved.

Acknowledgments. The author is very much grateful to A. Suslin, M. Ojanguren, J.-L. Colliot-Thélène, J.-P. Serre, Ph. Gille, N. Vavilov, A. Stavrova, V. Chernousov, A. Merkurjev, R. Fedorov, K. Zainoulline, D. Orlov and many, many others for their deep and very stimulating interest in the topic of the present paper.

References

François Charles and Bjorn Poonen (2016). "Bertini irreducibility theorems over finite fields". *J. Amer. Math. Soc.* 29.1, pp. 81–94. arXiv: 1311.4960. MR: 3402695.

V. Chernousov (2010). "Variations on a theme of groups splitting by a quadratic extension and Grothendieck-Serre conjecture for group schemes F_4 with trivial g_3 invariant". *Doc. Math.* Extra vol.: Andrei A. Suslin sixtieth birthday, pp. 147–169. MR: 2804252 (cit. on p. 230).

Jean-Louis Colliot-Thélène and Manuel Ojanguren (1992). "Espaces principaux homogènes localement triviaux". *Inst. Hautes Études Sci. Publ. Math.* 75, pp. 97–122. MR: 1179077 (cit. on pp. 222, 228, 229).

Jean-Louis Colliot-Thélène and Jean-Jacques Sansuc (1987). "Principal homogeneous spaces under flasque tori: applications". *J. Algebra* 106.1, pp. 148–205. MR: 878473 (cit. on pp. 228, 229).

M. Demazure and A. Grothendieck (1970a). *Schémas en groupes. II: Groupes de type multiplicatif, et structure des schémas en groupes généraux*. Séminaire de Géométrie Algébrique du Bois Marie 1962/64 (SGA 3). Dirigé par M. Demazure et A. Grothendieck. Lecture Notes in Mathematics, Vol. 152. Springer-Verlag, Berlin-New York, pp. ix+654. MR: 0274459 (cit. on pp. 231, 233, 235, 237).

– (1970b). *Schémas en groupes. III: Structure des schémas en groupes réductifs*. Séminaire de Géométrie Algébrique du Bois Marie 1962/64 (SGA 3). Dirigé par M. Demazure et A. Grothendieck. Lecture Notes in Mathematics, Vol. 153. Springer-Verlag, Berlin-New York, pp. viii+529. MR: 0274460 (cit. on pp. 225, 227).

Roman Fedorov (Jan. 2015). "On the Grothendieck-Serre conjecture on principal bundles in mixed characteristic". arXiv: 1501.04224 (cit. on p. 230).

– (2016). "Affine Grassmannians of group schemes and exotic principal bundles over $\mathbb{A}^1$". *Amer. J. Math.* 138.4, pp. 879–906. MR: 3538146 (cit. on p. 231).

Roman Fedorov and Ivan Panin (2015). "A proof of the Grothendieck-Serre conjecture on principal bundles over regular local rings containing infinite fields". *Publ. Math. Inst.*

Hautes Études Sci. 122, pp. 169–193. MR: 3415067 (cit. on pp. 222, 226, 227, 230, 231, 233–235).

O. Gabber (n.d.). *Announced and still unpublished*.

P. Gille (2002). "Torseurs sur la droite affine". *Transform. Groups* 7.3, pp. 231–245. MR: 1923972.

Philippe Gille (2009). "Le problème de Kneser-Tits". *Astérisque* 326. Séminaire Bourbaki. Vol. 2007/2008, Exp. No. 983, vii, 39–81 (2010). MR: 2605318.

A Grothendieck (1968a). *Le group de Brauer II*. Amsterdam: North-Holland (cit. on pp. 227, 229).

A. Grothendieck (1958). *Sur les faisceaux algébriques et les faisceaux analytiques cohérents*. Vol. 16. Séminaire Henri Cartan; 9e année: 1956/57. Quelques questions de topologie, Exposé no. 2. Secretariat mathématique, Paris, 73 pp. (mimeographed). MR: 0106907 (cit. on pp. 227, 229).

– (1961). "Éléments de géométrie algébrique. III. Étude cohomologique des faisceaux cohérents. I". *Inst. Hautes Études Sci. Publ. Math.* 11, p. 167. MR: 0163910.

– (1965). "Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. II". *Inst. Hautes Études Sci. Publ. Math.* 24, p. 231. MR: 0199181 (cit. on p. 229).

Alexander Grothendieck (1968b). "Le groupe de Brauer. III. Exemples et compléments". In: *Dix exposés sur la cohomologie des schémas*. Vol. 3. Adv. Stud. Pure Math. North-Holland, Amsterdam, pp. 88–188. MR: 244271.

Günter Harder (1968). "Halbeinfache Gruppenschemata über vollständigen Kurven". *Invent. Math.* 6, pp. 107–149. MR: 0263826 (cit. on pp. 234, 237).

M.-A. Knus, Raman Parimala, and R. Sridharan (1981/82). "Nonfree projective modules over $\mathbf{H}[X, Y]$ and stable bundles over $P_2(\mathbf{C})$". *Invent. Math.* 65.1, pp. 13–27. MR: 636877 (cit. on p. 231).

Fabien Morel and Vladimir Voevodsky (1999). "$\mathbf{A}^1$-homotopy theory of schemes". *Inst. Hautes Études Sci. Publ. Math.* 90, 45–143 (2001). MR: 1813224 (cit. on p. 232).

E. A. Nisnevič (1977). "Affine homogeneous spaces and finite subgroups of arithmetic groups over function fields". *Funkcional. Anal. i Priložen.* 11.1, pp. 73–74. MR: 0476772 (cit. on pp. 228, 229).

Yevsey A. Nisnevich (1984). "Espaces homogènes principaux rationnellement triviaux et arithmétique des schémas en groupes réductifs sur les anneaux de Dedekind". *C. R. Acad. Sci. Paris Sér. I Math.* 299.1, pp. 5–8. MR: 756297 (cit. on p. 229).

M. Ojanguren, I. Panin, and K. Zainoulline (2004). "On the norm principle for quadratic forms". *J. Ramanujan Math. Soc.* 19.4, pp. 289–300. MR: 2125505 (cit. on pp. 228, 230).

M. Ojanguren, Raman Parimala, and R. Sridharan (1983). "Indecomposable quadratic bundles of rank $4n$ over the real affine plane". *Invent. Math.* 71.3, pp. 643–653. MR: 695911 (cit. on p. 231).

Manuel Ojanguren (1980). "Quadratic forms over regular rings". *J. Indian Math. Soc. (N.S.)* 44.1-4, 109–116 (1982). MR: 752647 (cit. on pp. 220, 229).

Manuel Ojanguren and Ivan Panin (1999). "A purity theorem for the Witt group". *Ann. Sci. École Norm. Sup. (4)* 32.1, pp. 71–86. MR: 1670591.

– (2001). "Rationally trivial Hermitian spaces are locally trivial". *Math. Z.* 237.1, pp. 181–198. MR: 1836777 (cit. on pp. 229, 233).

I. A. Panin (2016a). "On Grothendieck-Serre's conjecture concerning principal G-bundles over reductive group schemes: II". *Izv. Ross. Akad. Nauk Ser. Mat.* 80.4, pp. 131–162. MR: 3535361 (cit. on pp. 228, 230).

– (2016b). "On Grothendieck-Serre's conjecture concerning principal G-bundles over reductive group schemes: II". *Izv. Ross. Akad. Nauk Ser. Mat.* 80.4, pp. 131–162. MR: 3535361.

I. A. Panin and A. K. Stavrova (2016). "On the Grothendieck-Serre conjecture concerning principal G-bundles over semi-local Dedekind domains". *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 443.Voprosy Teorii Predstavleniĭ Algebr i Grupp. 29, pp. 133–146. arXiv: 1512.00354. MR: 3507770 (cit. on p. 230).

I. A. Panin and A. A. Suslin (1997). "On a conjecture of Grothendieck concerning Azumaya algebras". *Algebra i Analiz* 9.4, pp. 215–223. MR: 1604322 (cit. on pp. 228, 229).

I. Panin, V. Petrov, and A. Stavrova (May 2009). "Grothendieck-Serre conjecture for adjoint groups of types E_6 and E_7 and for certain classical groups". arXiv: 0905.1427 (cit. on p. 230).

I. Panin, A. Stavrova, and N. Vavilov (2015a). "On Grothendieck-Serre's conjecture concerning principal G-bundles over reductive group schemes: I". *Compos. Math.* 151.3, pp. 535–567. arXiv: 0905.1418. MR: 3320571 (cit. on p. 230).

– (2015b). "On Grothendieck-Serre's conjecture concerning principal G-bundles over reductive group schemes: I". *Compos. Math.* 151.3, pp. 535–567. MR: 3320571 (cit. on pp. 230–233).

Ivan Panin (2015). "Proof of Grothendieck–Serre conjecture on principal bundles over regular local rings containing a finite field" (cit. on p. 230).

– (July 2017a). "Nice triples and Grothendieck–Serre's conjecture concerning principal G-bundles over reductive group schemes". arXiv: 1707.01756 (cit. on p. 236).

– (July 2017b). "Proof of Grothendieck–Serre conjecture on principal bundles over regular local rings containing a finite field". arXiv: 1707.01767 (cit. on pp. 227, 230, 236, 237).

– (July 2017c). "Two purity theorems and the Grothendieck–Serre's conjecture concerning principal G-bundles". arXiv: 1707.01763 (cit. on p. 228).

S. Parimala (1978). "Failure of a quadratic analogue of Serre's conjecture". *Amer. J. Math.* 100.5, pp. 913–924. MR: 517136 (cit. on p. 231).

Victor Petrov and Anastasia Stavrova (Nov. 2009). "Grothendieck-Serre conjecture for groups of type F_4 with trivial f_3 invariant". arXiv: 0911.3132 (cit. on p. 230).

Bjorn Poonen (2004). "Bertini theorems over finite fields". *Ann. of Math. (2)* 160.3, pp. 1099–1127. MR: 2144974 (cit. on p. 237).

Dorin Popescu (1986). "General Néron desingularization and approximation". *Nagoya Math. J.* 104, pp. 85–115. MR: 868439.

M. S. Raghunathan (1989). "Principal bundles on affine space and bundles on the projective line". *Math. Ann.* 285.2, pp. 309–332. MR: 1016097.

– (1994). "Principal bundles admitting a rational section". *Invent. Math.* 116.1-3, pp. 409–423. MR: 1253199 (cit. on p. 229).

– (1995). "Erratum: "Principal bundles admitting a rational section" [Invent. Math. 116 (1994), no. 1-3, 409–423.]" *Invent. Math.* 121.1, p. 223. MR: 1345290 (cit. on p. 229).

M. S. Raghunathan and A. Ramanathan (1984). "Principal bundles on the affine line". *Proc. Indian Acad. Sci. Math. Sci.* 93.2-3, pp. 137–145. MR: 813075.

J.-P. Serre (1958). "Espaces fibrés algébriques". In: *Anneaux de Chow et applications, Séminaire Chevalley, 2-e année*. Secrétariat mathématique, Paris (cit. on pp. 220, 227–229).

Richard G. Swan (1998). "Néron-Popescu desingularization". In: *Algebra and geometry (Taipei, 1995)*. Vol. 2. Lect. Algebra Geom. Int. Press, Cambridge, MA, pp. 135–192. MR: 1697953.

K. V. Zaĭnullin (2000). "On Grothendieck's conjecture on principal homogeneous spaces for some classical algebraic groups". *Algebra i Analiz* 12.1, pp. 150–184. MR: 1758566 (cit. on pp. 228, 229).

Received 2017-11-16.

Ivan Panin
St. Petersburg Branch of V. A. Steklov Mathematical Institute
Fontanka 27
191023 St. Petersburg
Russia

and

St. Petersburg State University
Department of Mathematics and Mechanics
Universitetsky prospekt, 28
198504, Peterhof, St. Petersburg
Russia

paniniv@gmail.com

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (241–266)

REPRESENTATIONS OF FINITE GROUPS AND APPLICATIONS

PHAM HUU TIEP

Abstract

We discuss some basic problems in representation theory of finite groups, and current approaches and recent progress on some of these problems. We will also outline some applications of these and other results in representation theory of finite groups to various problems in group theory, number theory, and algebraic geometry.

1 Introduction

Let G be a finite group and $\mathbb{F}$ be a field. A (finite-dimensional) *representation of G over* $\mathbb{F}$ is a group homomorphism $\Phi : G \to \mathrm{GL}(V)$ for some finite-dimensional vector space V over $\mathbb{F}$. Such a representation Φ is called *irreducible* if $\{0\}$ and V are the only $\Phi(G)$-invariant subspaces of V.

Representation theory of finite groups started with the letter correspondence between Richard Dedekind and Ferdinand Georg Frobenius in April (12th, 17th, 26th) 1896. In the same year, Frobenius constructed the character table of $\mathrm{PSL}_2(p)$, p any prime. Later on, the foundations of the *complex* representation theory (i.e. when $\mathbb{F} = \mathbb{C}$), were developed by Frobenius, Dedekind, Burnside, Schur, Noether, and others. Foundations of the *modular* representation theory (that is, when $p = \mathrm{char}(\mathbb{F}) > 0$ and p divides $|G|$) were (almost singlehandedly) laid out by Richard Brauer, started in 1935 and continued over the next few decades.

A natural question arises: in a more-than-century-old theory such as the representation theory of finite groups, what could still remain to be studied? By the Jordan–Hölder theorem, irreducible representations are the building blocks of any finite-dimensional representation of any finite group G. The main problem of representation theory of finite groups, which still remains wide open in full generality as well as for many important families of finite groups, can be formulated as follows:

The author gratefully acknowledges the support of the NSF (grants DMS-1201374 and DMS-1665014), the Simons Foundation, and the Clay Mathematics Institute.

MSC2010: primary 20C15; secondary 20C20, 20C33.

Problem 1.1. *Given a finite group G and a field $\mathbb{F}$, describe all irreducible representations of G over $\mathbb{F}$.*

Likewise, finite *simple* groups are building blocks of any finite group, and they are known thanks to the *Classification of Finite Simple Groups* (CFSG) Gorenstein, Lyons, and Solomon [1994], arguably one of the most monumental achievements of modern mathematics. So it is natural to focus our attention on studying Problem 1.1 for groups G that are simple, or more generally, *almost quasisimple*, that is, when $S \lhd G/\mathbf{Z}(G) \leq \mathrm{Aut}(S)$ for some finite non-abelian simple group S. Aside from the symmetric group S_n and alternating group A_n of degree n, almost quasisimple groups include the finite classical groups with natural module $V = \mathbb{F}_q^n$ (such as the special linear group $\mathrm{SL}(V) \cong \mathrm{SL}_n(q)$, the special unitary group $\mathrm{SU}(V) \cong \mathrm{SU}_n(q^{1/2})$ when q is a square, the symplectic group $\mathrm{Sp}(V) \cong \mathrm{Sp}_n(q)$ when $2|n$, and the special orthogonal groups $\mathrm{SO}(V)$), as well as their exceptional and twisted analogues. When q is a power of a fixed prime p, the latter are usually referred to as *finite groups of Lie type in characteristic p*. A more precise definition and a technically convenient framework, particularly for the Deligne-Lusztig theory Lusztig [1988, 1984], are provided by viewing the latter groups as the fixed point subgroups

$$\mathcal{G}^F := \{g \in \mathcal{G} \mid F(g) = g\}$$

for a Steinberg endomorphism $F : \mathcal{G} \to \mathcal{G}$ on a connected reductive algebraic group $\mathcal{G}$ defined over a field of characteristic p.

Throughout the paper, for a finite group G, $\mathrm{Irr}(G)$ denotes the set of complex irreducible characters of G, and $\mathrm{IBr}_p(G)$ denotes the set of irreducible p-Brauer characters of G (that is, the Brauer characters of irreducible representations of G over $\overline{\mathbb{F}}_p$) for a given prime p.

Example 1.2. Just to see how difficult Problem 1.1 can be, let us consider the example of the symmetric group $G = \mathsf{S}_n$.

(i) When $\mathbb{F} = \mathbb{C}$, according to the classical theory of Frobenius and Young, the complex irreducible characters $\chi = \chi^\lambda$ of G are labeled by the partitions λ of n, and the *hook length formula* gives us the degree $\chi^\lambda(1)$ for any $\lambda \vdash n$, and the *Frobenius character formula* determines the character values $\chi^\lambda(g)$ for all $g \in G$. Nevertheless, for a random partition λ of a large n and a random permutation $g \in \mathsf{S}_n$, it remains a difficult problem to compute $\chi^\lambda(g)$ efficiently. Even by now one still does not know a precise formula for the largest degree $b(\mathsf{S}_n) = \max_{\lambda \vdash n} \chi^\lambda(1)$ (which has importance in probabilistic group theory and various applications). The best until now, still asymptotic, answer to this question is given by work of Vershik and Kerov [1985] and independently of Logan and Shepp [1977] in 1977:

$$e^{-1.2826\sqrt{n}}\sqrt{n!} \leq b(\mathsf{S}_n) \leq e^{-0.1156\sqrt{n}}\sqrt{n!}.$$

(This result implies that, in a sense, a randomly chosen partition λ of n already yields an irreducible character of degree close to be largest possible, and thus explains the difficulty of the question.)

(ii) For various applications, one also needs to know good *exponential* bounds on character values for symmetric groups. For instance, for each $1 \neq g \in \mathsf{S}_n$ one would like to find an (explicit) constant $0 < \alpha = \alpha(g) < 1$ such that $|\chi(g)| \leq \chi(1)^{\alpha}$ for all $\chi \in \mathrm{Irr}(\mathsf{S}_n)$. Such an $\alpha(g)$ was found by Fomin and Lulov [1995] in the case all cycles of g have same size. The general case was settled by Larsen and Shalev [2008] only in 2008, and it led to important results in a number of applications.

(iii) Now let us keep the same group $G = \mathsf{S}_n$ but change $\mathbb{F}$ to $\mathbb{F}_2$. Then the irreducible representations of G over $\mathbb{F}$ are labeled by partitions of λ into distinct parts. But now, for a given degree, say $n = 1000$, and given such a partition λ, one still does not know what is the dimension of the corresponding representation. The same story goes with the similar question for $G = \mathrm{GL}_{1000}(2)$ and $\mathbb{F} = \mathbb{F}_2$ or $\mathbb{F} = \mathbb{F}_3$.

Various questions mentioned in Example 1.2 also remain open, say, for most of the finite groups of Lie type.

Problem 1.3. *Let G be a finite group of Lie type. For any $g \in G \smallsetminus \mathbf{Z}(G)$, find a constant $0 < \alpha = \alpha(g) < 1$, as small and explicit as possible, such that $|\chi(g)| \leq \chi(1)^{\alpha}$ for all $\chi \in \mathrm{Irr}(G)$.*

Given a finite group G, let $\mathfrak{d}_p(G)$ denote the smallest degree of *faithful* representations of G over $\overline{\mathbb{F}}_p$. We would like to study the following special instance of Problem 1.1, which turns out to be of importance for many applications:

Problem 1.4. *Given an almost quasisimple group G and a prime p,*

(i) *determine $\mathfrak{d}_p(G)$, and*

(ii) *classify irreducible $\overline{\mathbb{F}}_p G$-representations of degree up to $\mathfrak{d}_p(G)^{2-\epsilon}$, for a fixed $0 < \epsilon < 1$.*

To formulate further conjectures, let us introduce some more notation. For a fixed group G and p, let $P \in \mathrm{Syl}_p(G)$, and let

$$\mathrm{Irr}_{p'}(G) = \{\chi \in \mathrm{Irr}(G) \mid p \nmid \chi(1)\}.$$

We say that $\chi \in \mathrm{Irr}(G)$ has *p-defect* 0 if $\chi(1)_p = |G|_p$. If χ belongs to a p-block B of G with defect group D, then χ is said to have *height* 0 if $\chi(1)_p = [G : D]_p$.

Several fundamental conjectures in representation theory of finite groups follow the *global-local principle*, which in this case states that *certain global invariants of a finite group G can be determined locally, in terms of its p-subgroups, their normalizers, etc.*

The following is probably the easiest one to formulate among all the global-local conjectures:

Conjecture 1.5 (McKay [1972]). *There exists a bijection* $\mathrm{Irr}_{p'}(G) \stackrel{\pi}{\longleftrightarrow} \mathrm{Irr}_{p'}(\mathbf{N}_G(P))$.

The *Alperin-McKay conjecture* Alperin [1976] is a blockwise version of the McKay conjecture and asserts: *If a p-block B of a finite group G has a defect group D and Brauer correspondent b, a p-block of $\mathbf{N}_G(D)$, then B and b have the same number of characters of height* 0. There are several recent refinements (due to Isaacs, Navarro, Turull, and others) of the McKay conjecture, which roughly say that in Conjecture 1.5 there should exist a bijection π that is compatible with the action of certain Galois automorphisms of $\overline{\mathbb{Q}}$ and preserving congruences modulo p, local Schur indices, etc.

Even if the Problem 1.1 remains unsolved, can one hope for a "natural" labeling of the irreducible representations of G? If G is a connected reductive algebraic group defined over $\mathbb{F}$, then one can label the finite-dimensional rational irreducible representations of G by their highest weights. Alperin [1987] conjectured in 1986 that one should be able to do the same for any finite group G. More precisely, a *p-weight* of G is a pair (Q, δ), where Q is a p-subgroup of G and $\delta \in \mathrm{Irr}(\mathbf{N}_G(Q)/Q)$ has p-defect 0.

Conjecture 1.6 (**Alperin**). *The number of irreducible p-Brauer characters of a finite group G equals the number of G-conjugate classes of p-weights of G.*

Given a p-block B of G, a *p-weight of B* is a p-weight (Q, δ) with δ belonging to an $\mathbf{N}_G(Q)$-block b with $b^G = B$. Then the blockwise version of the Alperin weight Conjecture 1.6 asserts that *the number of irreducible p-Brauer characters of G that belong to B equals the number of G-conjugate classes of p-weights of B.*

Finally, we recall the *Brauer height zero conjecture* Brauer [1956], perhaps one of the oldest and deepest conjectures in the modular representation theory:

Conjecture 1.7 (**Brauer**, 1955). *All complex irreducible characters in a p-block B of a finite group G have height zero if and only if the defect groups D of B are abelian.*

2 Current Approaches and Recent Results

The aforementioned, and several other, fundamental conjectures in representation theory of finite groups have been proved to hold for many classes of finite groups, including solvable groups, p-solvable groups, as well as various families of simple groups. As important evidence in favor of these conjectures as these results are, none of the above conjectures has been proved to hold for *all* arbitrary finite groups.

A possible approach to tackle these conjectures, which was already one of the main ideas in Dade's works in the '90s, is to try to use the CFSG to reduce to *simple* groups. As we will see, such reductions are possible, on the one hand, and they have led to important progress on some of these conjectures. On the other hand, oftentimes such reductions require one to prove much stronger statements about the simple groups, not merely the original conjecture in question. We will now discuss recent progress on various problems mentioned in §1.

2.1 The McKay conjecture.

In 2007, Isaacs, Malle, and Navarro succeeded in proving the following reduction theorem for the McKay Conjecture 1.5:

Theorem 2.1.1. Isaacs, Malle, and Navarro [2007] *Suppose that every finite non-abelian simple group S is* **McKay-good** *for the prime p. Then the McKay conjecture holds for arbitrary finite groups (for the prime p).*

Here, the *McKay-goodness* (also known as *the inductive McKay condition*) for the prime p is much more than just satisfying the McKay conjecture. It is in fact a long and complicated list of conditions concerning representations and cohomology of certain subgroups of the universal cover of S, occupying a couple of pages of Isaacs, Malle, and Navarro [ibid.]. Later, a reduction theorem in the same spirit for the Alperin-McKay conjecture was obtained by Späth in Späth [2013a]. Combined efforts of Malle, Cabanes, and Späth have also led to the proof of the inductive McKay condition for all simple groups, except for simple orthogonal groups in odd characteristics and exceptional groups of type E_6, ${}^2E_6$, and E_7. Moreover, a breakthough has recently been achieved by Malle and Späth, showcasing the strengths of the current approach:

Theorem 2.1.2. Malle and Späth [2016] *The McKay conjecture for $p = 2$ holds for all finite groups G.*

Less is currently known about various refinements of the McKay conjecture, which, if true, would imply many interesting consequences. For instance, the Galois-McKay conjecture, as proposed by Navarro in Navarro [2004], implies that *the character table of any finite group G detects whether a Sylow p-subgroup of G is self-normalizing*. This would give a partial answer to Problem 12 in Brauer's celebrated list Brauer [1963]: *Given the character table of a group G and a prime p dividing $|G|$, how much information about the Sylow p-groups P of G can be obtained?* In fact, a unconditional answer has been obtained, supporting the Galois-McKay conjecture:

Theorem 2.1.3. *Let G be a finite group, p a prime, and let $P \in \mathrm{Syl}_p(G)$.*

(i) *Navarro, Tiep, and Turull [2007] Suppose $p > 2$. Then $P = \mathbf{N}_G(P)$ if and only if $\mathrm{Irr}_{p'}(G)$ contains a unique character χ with $\mathbb{Q}(\chi) \subseteq \mathbb{Q}(\exp(2\pi i/|G|_{p'}))$.*

(ii) *Schaeffer Fry [2016] and Schaeffer Fry and J. Taylor [2018] Suppose $p = 2$ and let σ be the automorphism of $\mathbb{Q}(\exp(2\pi i/|G|))$ that fixes every root of unity of 2-power order and squares every root of unity of odd order. Then $P = \mathbf{N}_G(P)$ if and only if every $\chi \in \mathrm{Irr}_{2'}(G)$ is fixed by σ.*

Finite groups with self-normalizing Sylow p-subgroups (with $p > 2$) also stand out as one of the few cases where a *canonical* bijection $\boldsymbol{\pi}$ satisfying the McKay Conjecture 1.5 can be found (and therefore it is compatible with the action of Galois automorphisms). As shown in Navarro, Tiep, and Vallejo [2014], in this case $\boldsymbol{\pi}(\chi)$ can be taken to be the (unique) linear constituent of $\chi|_P$ for any $\chi \in \mathrm{Irr}_{p'}(G)$. (Also see Isaacs [1973], Navarro [2003], Giannelli, Kleshchev, Navarro, and Tiep [2017], Isaacs, Navarro, Olsson, and Tiep [2017], Giannelli, Tent, and Tiep [2018] for some other occurrences of canonical McKay correspondences.) We also note the following recent result:

Theorem 2.1.4. Guralnick, Navarro, and Tiep [2016] *Let G be a finite group, p be a prime, and $P \in \mathrm{Syl}_p(G)$. Suppose that $\mathbf{N}_G(P)$ has odd order. Then the McKay conjecture, the Alperin weight conjecture, and their blockwise versions hold for G and the prime p.*

2.2 The Alperin weight conjecture (AWC).

The following reduction theorem for the Alperin weight Conjecture 1.6 was proved in 2011:

Theorem 2.2.1. *Navarro and Tiep [2011] Suppose that every finite non-abelian simple group S is* **AWC-good** *for the prime p. Then the Alperin weight conjecture holds for arbitrary finite groups (for the prime p).*

Another reduction theorem for the AWC was obtained by Puig [2011], and the blockwise version of the AWC was reduced to simple groups by Späth in Späth [2013b].

As it was the case with the inductive McKay condition, the *AWC-goodness* in Theorem 2.2.1 is much stronger than just satisfying the AWC. Nevertheless, the list of AWC-good simple groups for the prime p has been shown to include the simple groups of Lie type in the same characteristic p, the alternating groups, and all the sporadic simple groups.

2.3 The Brauer height zero conjecture (BHZ).

The "if" direction of the Brauer height zero Conjecture 1.7 was reduced by Berger and Knörr [1988] in 1988 to quasisimple groups. This was a "pure" reduction; namely they showed that *if the "if" direction of the BHZ holds for all blocks (with abelian defect groups) of all finite quasisimple groups, then it also holds for all blocks (with abelian defect groups) of all finite groups.* The verification of (both directions of) the BHZ for quasisimple groups was completed recently by Kessar and Malle [2013, 2017]. Thus *the "if" direction of the BHZ holds for arbitrary finite groups.*

The “only if” direction of the BHZ is even more difficult. The Gluck–Wolf proof of this direction for p-solvable groups was already extraordinarily complicated. For arbitrary finite groups, the following reduction theorem was obtained in 2014:

Theorem 2.3.1. *Navarro and Späth [2014] Suppose all of the following statements hold:*

(i) *The inductive Alperin-McKay condition Späth [2013a] holds for all finite simple groups S for the prime p;*

(ii) *A generalized Gluck-Wolf theorem (gGW) holds; and*

(iii) *The “only if” direction of the BHZ holds for all finite quasisimple groups.*

Then the “only if” direction of the BHZ holds for all finite groups for the prime p.

Another reduction theorem, with condition 2.3.1(i) replaced by the projective Dade conjecture Dade [1994], was obtained earlier by Murai in Murai [2012]. As mentioned above, condition (iii) in Theorem 2.3.1 holds, thanks to Kessar and Malle [2017]. The statement (gGW) alluded to in Theorem 2.3.1(ii) is a relative version of the BHZ, and is now also a theorem:

Theorem 2.3.2. *Navarro and Tiep [2013] Let G be a finite group with a normal subgroup Z, p be a prime, and let $\lambda \in \mathrm{Irr}(Z)$. Suppose that $\chi(1)/\lambda(1)$ is coprime to p for all $\chi \in \mathrm{Irr}(G)$ lying above λ. Then the Sylow p-subgroups of G/Z are abelian.*

Let us also mention

Theorem 2.3.3. *Navarro and Tiep [2012] The Brauer height zero conjecture holds for all 2-blocks of G with defect groups $P \in \mathrm{Syl}_2(G)$.*

2.4 Dimensions of irreducible representations and Problem 1.4. For an almost quasisimple group G, let S denote the unique non-abelian composition factor of G. In the case S is a sporadic simple group, Problem 1.4 depends largely on latest developments in computational group theory. In particular, $\mathfrak{d}_p(G)$ has been completely determined Jansen [2005]. However, Problem 1.4 remains open for a number of large sporadic groups, for instance in the case where $G = M$ is *the Monster* and $p = 2$.

Next, let us consider the case $S = \mathsf{A}_n$ with $n \geq 5$. Here, certainly the case of complex representations is very well understood, thanks to classical work of Frobenius and Schur. The case of modular representations of S_n was settled by James in James [1983]. In particular, he showed that

$$\mathfrak{d}_p(\mathsf{S}_n) = \begin{cases} n-1, & p \nmid n \\ n-2, & p \mid n \end{cases} .$$

In fact, James proved that, for a fixed p-regular partition $\mu = (\lambda_2, \dots, \lambda_k)$ of m,

$$\dim D^{(n-m,\lambda_2,\dots,\lambda_k)} \approx \frac{n^m}{m!} \dim D^{\mu}$$

when $n \to \infty$ (if D^λ is the p-modular irreducible representation of S_n labeled by the p-regular partition $\lambda \vdash n$). This beautiful result gives however only an asymptotic bound on the dimension of D^λ. For a number of applications, one needs an effective bound on $\dim D^\lambda$, and the first such bound was obtained in Guralnick, Larsen, and Tiep [2012]. Define for $p \neq 2$

$$m_p(\lambda) := \max(\lambda_1, (\lambda^{\mathbf{M}})_1),$$

(the longest row of partitions λ and $\lambda^{\mathbf{M}}$), where $\lambda \mapsto \lambda^{\mathbf{M}}$ is the Mullineux bijection on the set of p-regular partitions of n; also set $m_2(\lambda) := \lambda_1$.

Theorem 2.4.1. *Guralnick, Larsen, and Tiep [ibid.] For any $p \geq 0$, and any p-regular partition λ of n,*

$$\dim D^\lambda \geq 2^{\frac{n-m_p(\lambda)}{2}}.$$

This effective bound was used to establish polynomial representation growth for the modular representations of S_n and A_n, see Guralnick, Larsen, and Tiep [ibid., Theorem 1.1]. It also allowed to deduce quantitative results on branching rules for irreducible S_n-representations over A_n, and, as a consequence, resolve Problem 1.4 for $G = \mathsf{A}_n$. However, the bound in Theorem 2.4.1 is not of the right magnitude. This issue has been rectified very recently:

Theorem 2.4.2. *Kleshchev and Tiep [n.d.] Let p be a prime, $m \geq 2$, and let $\lambda = (n - m, \lambda_2, \dots, \lambda_k)$ be a p-regular partition of $n \geq (m-1)p + 2$. Then*

$$\dim D^\lambda \geq \begin{cases} \left(\prod_{j=0}^{m-1}(n - jp)\right)/m!, & p \geq 5 \\ \left(\prod_{j=0}^{m-1}(n-2-jp)\right)/m!, & p = 2, 3 \end{cases}.$$

For *spin* representations, i.e. faithful representations of double covers of S_n and A_n, Problem 1.4 has been studied in Kleshchev and Tiep [2004, 2012]. In particular, it was shown in Kleshchev and Tiep [2004] that

$$\mathfrak{d}_p(2\mathsf{S}_n) = \begin{cases} 2^{\lfloor (n-1)/2 \rfloor}, & p \nmid n \\ 2^{\lfloor (n-2)/2 \rfloor}, & p \mid n. \end{cases}$$

Furthermore, irreducible spin modular representations of degree up to $(n/2) \cdot \mathfrak{d}_p(G)$ for $G = 2\mathsf{S}_n$ and $2\mathsf{A}_n$ were determined in Kleshchev and Tiep [2012].

Now we discuss the main case of Problem 1.4 when S is a simple group of Lie type in characteristic ℓ. In the defining characteristic case, that is when $p = \mathrm{char}(\mathbb{F}) = \ell$, Problem 1.4 can be solved using the representation theory of reductive algebraic groups (namely the theory of highest weight modules), and Premet's theorem Premet [1987]. In fact, this was done by Liebeck for classical groups, and by Lübeck for exceptional groups.

Next we consider the cross characteristic case, that is when $p = \mathrm{char}(\mathbb{F}) \neq \ell$. If, moreover, $p = 0$ or $0 < p \nmid |G|$, then Problem 1.4 can be solved using the *Deligne-Lusztig theory* Lusztig [1988, 1984]. This was done in Tiep and Zalesskii [1996] for classical groups, and by Lübeck for exceptional groups.

The remaining case ($\ell \neq p > 0$ and $p||G|$) turns out to be much harder and is still ongoing. Complete results have been obtained for groups of type A, that is when $G = \mathrm{SL}_n(q)$, see Guralnick and Tiep [1999] and Brundan and Kleshchev [2000], and when $G = \mathrm{SU}_n(q)$, see Hiss and Malle [2001] and Guralnick, Magaard, Saxl, and Tiep [2002].

Theorem 2.4.3. *Guralnick and Tiep [1999] Assume $n \geq 4$ and $(n,q) \neq (4,2), (4,3)$. Then*

$$\mathfrak{d}_p(\mathrm{SL}_n(q)) = \frac{q^n-1}{q-1} - \begin{cases} 1, & p \nmid \frac{q^n-1}{q-1} \\ 2, & p | \frac{q^n-1}{q-1}. \end{cases}$$

Moreover, $\mathrm{SL}_n(q)$ *has one irreducible representation over* $\mathbb{F}$ *of degree* $\mathfrak{d}_p$ *and* $(q-1)_{p'}-1$ *of degree* $(q^n-1)/(q-1)$*. All other nontrivial irreducible representations have degree at least* $(q^{n-1}-1)\left(\dfrac{q^{n-2}-q}{q-1}-1\right)$.

Theorem 2.4.4. *Guralnick, Magaard, Saxl, and Tiep [2002] Assume $n \geq 5$ and $(n,q) \neq (6,2)$. Then*

$$\mathfrak{d}_p(\mathrm{SU}_n(q)) = \left\lfloor \frac{q^n-1}{q+1} \right\rfloor.$$

Moreover, $\mathrm{SU}_n(q)$ *has* $(q+1)_{p'}$ *irreducible representations over* $\mathbb{F}$ *of degree* $\mathfrak{d}_p$ *or* $\mathfrak{d}_p+1$*. All other nontrivial irreducible representations have degree at least* $\dfrac{(q^n-1)(q^{n-1}-q^2)}{(q+1)(q^2-1)}$.

The case of symplectic groups was also settled in Guralnick, Magaard, Saxl, and Tiep [ibid.] and Guralnick and Tiep [2004]; in particular,

$$\mathfrak{d}_p(Sp_{2n}(q)) = \begin{cases} (q^n-1)/2, & q \text{ odd} \\ (q^n-1)(q^n-q)/2(q+1), & q \text{ even.} \end{cases}$$

Less complete results have also been obtained for other families of finite groups of Lie type, see Tiep [2003], Tiep [2006] for relevant references. All these results rely crucially on the Deligne-Lusztig theory and further important results of Broué and Michel [1989] and Bonnafé and Rouquier [2003].

As mentioned in Example 1.2, of interest is also the *largest degree*

$$b(G) = \max_{\chi \in \mathrm{Irr}(G)} \chi(1)$$

of complex irreducible characters of an almost quasisimple group G. In the case G is of Lie type in characteristic p, the Steinberg character St has quite a large degree, equal to $|G|_p$. It turns out that $b(G)/\mathsf{St}(1)$ can grow unbounded when we fix the size q of the defining field $\mathbb{F}_q$ and let the rank r of G grow. An upper bound for $b(G)$ was given in Seitz [1990, Theorem 2.1], which yields the exact value of $b(G)$ when q is large enough compared to r.

Theorem 2.4.5. *Larsen, Malle, and Tiep [2013] For any $1 > \varepsilon > 0$, there are (explicit) constants $A, B > 0$ depending on ε such that, for any simple algebraic group $\mathcal{G}$ in characteristic p of rank r and any Steinberg endomorphism $F : \mathcal{G} \to \mathcal{G}$, the largest degree $b(G)$ of the corresponding finite group $G := \mathcal{G}^F$ over $\mathbb{F}_q$ satisfies the following inequalities:*

$$A(\log_q r)^{(1-\varepsilon)/\gamma} < \frac{b(G)}{|G|_p} < B(1 + \log_q r)^{2.54/\gamma}$$

if G is classical, and $1 \leq b(G)/|G|_p < B$ if G is an exceptional group of Lie type. Here, $\gamma = 1$ if G is untwisted of type A, and $\gamma = 2$ otherwise.

A lower bound for the largest degree of modular irreducible representations of $G = \mathcal{G}^F$ was also given in Larsen, Malle, and Tiep [ibid., Theorem 1.4]. Theorem 2.4.5 implies the following, somewhat surprising, consequence which answers a question raised by D. Vogan and J. Berstein:

Corollary 2.4.6. *Let q be any prime power and let $n > 2q^{6815}$. Consider a non-degenerate quadratic space $V = \mathbb{F}_q^n$, a non-degenerate subspace U of codimension 1 in V, and embed the full orthogonal group $H = \mathrm{GO}(U) \cong \mathrm{GO}^{\pm}_{n-1}(q)$ via $g \mapsto \mathrm{diag}(g, \det(g))$ in the special orthogonal group $G = \mathrm{SO}(V) \cong \mathrm{SO}^{\pm}_n(q)$. Then there exists a character $\chi \in \mathrm{Irr}(G)$ such that its restriction to H is not multiplicity-free (and χ is trivial at $\mathbf{Z}(G)$).*

Proof. We give a proof for the case $n = 2m + 1 \geq 9$; the case $n = 2m$ is completely similar. By Larsen, Malle, and Tiep [ibid., Theorem 5.2] (and its proof), there is $\chi \in \mathrm{Irr}(G)$ such that χ is trivial at $\mathbf{Z}(G)$ and

$$\chi(1) > q^{m^2} \cdot \frac{1}{5}\left(\log_q(m + 10)\right)^{3/8}.$$

On the other hand, $|\,\mathrm{Irr}(H)| < 15q^m$ by Theorems 3.14 and 3.21 of Fulman and Guralnick [2012]. Now if $\chi|_H$ is multiplicity-free, then by Schwarz's inequality we must have that

$$\chi(1) \leq \sum_{\alpha \in \mathrm{Irr}(H)} \alpha(1) \leq (|H| \cdot |\,\mathrm{Irr}(H)|)^{1/2} < \sqrt{30} \cdot q^{m^2},$$

contradicting the above lower bound on $\chi(1)$ when $m \geq q^{6815}$. □

In fact, Corollary 2.4.6 also holds if we take U to be any nonzero proper non-degenerate subspace of V and replace H with $\mathrm{SO}(V) \cap (\mathrm{GO}(U) \times \mathrm{GO}(U^{\perp}))$.

2.5 Bounds on character values: Problem 1.3. For a finite group G, a *character ratio* is a complex number of the form $\chi(g)/\chi(1)$, where $g \in G$ and χ is an irreducible character of G. Upper bounds for absolute values of character values and character ratios have long been of interest, for various reasons; these include applications to random generation, covering numbers, mixing times of random walks, word maps, representation varieties and other areas.

The first significant bounds on character ratios for finite groups of Lie type G, defined over a field $\mathbb{F}_q$, were obtained by Gluck [1993, 1997]. In particular, he showed that $|\chi(g)|/\chi(1) \leq Cq^{-1/2}$ for any non-central element $g \in G$ and any non-linear character $\chi \in \mathrm{Irr}(G)$, where C is an absolute constant. Another explicit character bound for finite classical groups, with natural module $V = \mathbb{F}_q^n$, was obtained in Larsen, Shalev, and Tiep [2011, Theorem 4.3.6]:

$$\frac{|\chi(g)|}{\chi(1)} < q^{-\sqrt{\mathsf{supp}(g)}/481},$$

where $\mathsf{supp}(g)$ is the codimension of the largest eigenspace of $g \in G$ on $V \otimes_{\mathbb{F}_q} \overline{\mathbb{F}}_q$.

These bounds have played a crucial role in a number of applications (some described in §3). However, in many situations of these and other applications, one needs stronger, exponential character bounds as described in Problem 1.3. Such a bound was established for S_n in Larsen and Shalev [2008]. For finite groups of Lie type, it has been obtained for the first time in Bezrukavnikov, Liebeck, Shalev, and Tiep [n.d.]. For a subgroup X of an algebraic group $\mathcal{G}$, write X_{unip} for the set of non-identity unipotent elements of X. For a fixed Steinberg endomorphism $F : \mathcal{G} \to \mathcal{G}$, a Levi subgroup $\mathcal{L}$ of $\mathcal{G}$ is called *split*, if it is an F-stable Levi subgroup of an F-stable proper parabolic subgroup of $\mathcal{G}$. For an F-stable Levi subgroup $\mathcal{L}$ of $\mathcal{G}$ and $L := \mathcal{L}^F$, we define

$$\alpha(L) := \max{}_{u \in L_{\mathrm{unip}}} \frac{\dim u^{\mathcal{L}}}{\dim u^{\mathcal{G}}}, \quad \alpha(\mathcal{L}) := \max{}_{u \in \mathcal{L}_{\mathrm{unip}}} \frac{\dim u^{\mathcal{L}}}{\dim u^{\mathcal{G}}}$$

if $\mathcal{L}$ is not a torus, and $\alpha(L) = \alpha(\mathcal{L}) := 0$ otherwise.

Theorem 2.5.1. *Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid.] There exists a function $f : \mathbb{N} \to \mathbb{N}$ such that the following statement holds. Let $\mathcal{G}$ be a connected reductive algebraic group such that $[\mathcal{G}, \mathcal{G}]$ is simple of rank r over a field of good characteristic $p > 0$. Let $G := \mathcal{G}^F$ for a Steinberg endomorphism $F : \mathcal{G} \to \mathcal{G}$. Let $g \in G$ be any* *element such that* $\mathbf{C}_G(g) \leq L := \mathcal{L}^F$, *where* $\mathcal{L}$ *is a split Levi subgroup of* $\mathcal{G}$. *Then, for any character* $\chi \in \mathrm{Irr}(G)$ *and* $\alpha := \alpha(L)$, *we have*

$$|\chi(g)| \leq f(r)\chi(1)^{\alpha}.$$

The α-bound in Theorem 2.5.1 is sharp in many cases; for instance, it is always optimal in the case $G = \mathrm{SL}_n(q)$, see Bezrukavnikov, Liebeck, Shalev, and Tiep [n.d., Theorem 1.3]. Furthermore, the function $f(r)$ is given explicitly in Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid., Proposition 2.7]. Explicit bounds for $\alpha(L)$ can be found in Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid.]; in particular, it is shown in Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid., Theorem 1.6] that

$$\alpha(L) \leq \alpha(\mathcal{L}) \leq \frac{1}{2}\left(1 + \frac{\dim \mathcal{L}}{\dim \mathcal{G}}\right)$$

if $\mathcal{G}$ is a classical group. As a consequence, the following Lie-theoretic analogue of the celebrated Fomin–Lulov bound Fomin and Lulov [1995] was obtained in Bezrukavnikov, Liebeck, Shalev, and Tiep [n.d.]:

Corollary 2.5.2. *Let* $m < n$ *be a divisor of* n *and let* $L \leq G = \mathrm{GL}_n(q)$ *be a Levi subgroup of the form* $L = \mathrm{GL}_{n/m}(q)^m$. *Let* $g \in G$ *with* $\mathbf{C}_G(g) \leq L$. *Then we have*

$$|\chi(g)| \leq f(n-1)\chi(1)^{\frac{1}{m}}$$

for all $\chi \in \mathrm{Irr}(G)$, *where* $f : \mathbb{N} \to \mathbb{N}$ *is the function specified in Theorem 2.5.1.*

Again, the exponent $1/m$ in Corollary 2.5.2 is sharp. Moreover, an exponential character bound for ℓ-Brauer characters of $G = \mathrm{GL}_n(q)$, $\mathrm{SL}_n(q)$ (in the case $\ell \nmid q$ and for the elements $g \in G$ with $\mathbf{C}_G(g)$ contained in a split Levi subgroup of G), has also been established in Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid.].

The above results do not cover, for instance, the case where $g \in \mathcal{G}^F$ is a unipotent element. However, a complete result covering all elements in $\mathrm{GL}_n(q)$ and $\mathrm{SL}_n(q)$ has been obtained in Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid.]:

Theorem 2.5.3. *There is a function* $h : \mathbb{N} \to \mathbb{N}$ *such that the following statement holds. For any* $n \geq 5$, *any prime power* q, *any irreducible complex character* χ *of* $G := \mathrm{GL}_n(q)$ *or* $\mathrm{SL}_n(q)$, *and any non-central element* $g \in G$,

$$|\chi(g)| \leq h(n) \cdot \chi(1)^{1-\frac{1}{2n}}.$$

There is also a different approach to Problem 1.3, which so far has been worked out completely for groups of type A in Guralnick, Larsen, and Tiep [n.d.]. It has long been observed that irreducible characters of symmetric groups and of finite groups of Lie type seem to appear in "clusters", where the characters in a given cluster have roughly the same degree (as a polynomial function of n in the case of S_n, and of q^r in the case of a Lie-type group of rank r over $\mathbb{F}_q$), and display roughly the same behavior in several contexts. The main goal of this approach is to develop the concept of *character level* – the characters $\chi \in \mathrm{Irr}(G)$ will then be grouped in clusters according to their level, and then prove exponential character bounds for characters (at least of not-too-large level).

Let us use the notation GL^ϵ to denote GL when $\epsilon = +$ and GU when $\epsilon = -$, and similarly for SL^ϵ. Let $V = \mathbb{F}_Q^n$ be the natural module of $G \in \{\mathrm{GL}_n^\epsilon(q), \mathrm{SL}_n^\epsilon(q)\}$, where $Q = q$ when $\epsilon = +$, and $Q = q^2$ when $\epsilon = -$. It is known that the class function

$$\tau : g \mapsto \epsilon^n (\epsilon q)^{\dim_{\mathbb{F}_Q} \mathrm{Ker}(g - 1_V)}$$

is a (reducible) character of G. The *true level* $\mathfrak{l}^*(\chi)$ of a character $\chi \in \mathrm{Irr}(G)$ is then defined to be the smallest non-negative integer j such that χ is an irreducible constituent of τ^j; and the *level* $\mathfrak{l}(\chi)$ is the smallest non-negative integer j such that $\lambda\chi$ is an irreducible constituent of τ^j for some character λ of degree 1 of G, see Guralnick, Larsen, and Tiep [ibid.].

Theorem 2.5.4. *Guralnick, Larsen, and Tiep [ibid.] Let $G \in \{\mathrm{GL}_n^\epsilon(q), \mathrm{SL}_n^\epsilon(q)\}$ with $n \geq 2$, $\epsilon = \pm$, and let $\chi \in \mathrm{Irr}(G)$ have level $j = \mathfrak{l}(\chi)$. Then the following statements hold.*

(i) $q^{j(n-j)}/2(q+1) \leq \chi(1) \leq q^{nj}$. *If furthermore $j \geq n/2$, then*

$$\chi(1) > q^{n^2/4-2}/(q-\epsilon).$$

(ii) *If $n \geq 7$ and $\lceil (1/n) \log_q \chi(1) \rceil < \sqrt{n-1} - 1$, then*

$$\mathfrak{l}(\chi) = \left\lceil \frac{\log_q \chi(1)}{n} \right\rceil.$$

(iii) *If $\mathfrak{l}(\chi) \leq \sqrt{(8n-17)/12} - 1/2$ for $\chi \in \mathrm{Irr}(G)$, then $|\chi(g)| < 2.43\chi(1)^{1-1/n}$ for all $g \in G \smallsetminus \mathbf{Z}(G)$. Moreover, if $\mathfrak{l}(\chi) \leq (\sqrt{12n-59} - 1)/6$ for $\chi \in \mathrm{Irr}(G)$, then*

$$|\chi(g)| < 2.43\chi(1)^{\max(1-1/2\mathfrak{l}(\chi), 1-\mathsf{supp}(g)/n)}$$

for all $g \in G$.

(iv) *Suppose that $g \in G$ satisfies $|\mathbf{C}_{\mathrm{GL}_n^\epsilon(q)}(g)| \leq q^{n^2/12}$. Then $|\chi(g)| \leq \chi(1)^{8/9}$.*

Note that the exponent $1 - 1/n$ in the character bound in Theorem 2.5.4(iii) is optimal. Another feature of character level is provided by the following result, which shows that the characters of level $j < n/2$ of $\mathrm{SL}_n^\epsilon(q)$ are controlled by $\mathrm{GL}_j^\epsilon(q)$:

Theorem 2.5.5. *Guralnick, Larsen, and Tiep [n.d.] For any $0 \leq j \leq n$, there is a canonical bijection $\alpha \mapsto \Theta(\alpha)$ between $\{\alpha \in \mathrm{Irr}(\mathrm{GL}_j^\epsilon(q)) \mid \mathfrak{l}^*(\alpha) \geq 2j - n\}$ and $\{\chi \in \mathrm{Irr}(\mathrm{GL}_n^\epsilon(q)) \mid \mathfrak{l}^*(\chi) = j\}$. If furthermore $j < n/2$, then the map $\alpha \mapsto \Theta(\alpha)|_S$ yields a canonical bijection between $\mathrm{Irr}(\mathrm{GL}_j^\epsilon(q))$ and $\{\theta \in \mathrm{Irr}(S) \mid \mathfrak{l}(\theta) = j\}$ for $S = \mathrm{SL}_n^\epsilon(q)$.*

3 Some Recent Applications

Results on Problems 1.3, 1.4 and other have been used in the *revision of the Classification of Finite Simple Groups* (for instance, in the classification of *quadratic modules*), in computational group theory (e.g. in the recognition of permutation/matrix groups of moderate degree). They also played a key role in the proofs Guralnick and Tiep [2005], Guralnick and Tiep [2008] of Larsen's conjecture on moments Katz [2004, 2005] and the Kollár-Larsen conjecture Balaji and Kollár [2008] on symmetric powers, and the solution Guralnick and Tiep [2012] of the Kollár-Larsen problem Kollár and Larsen [2009] on linear groups generated by elements of bounded deviation and crepant resolutions.

We will now discuss some recent applications in group theory, number theory, and algebraic geometry.

3.1 Automorphy lifting and adequate groups.

For a number field $\mathbb{K}$, let $G_{\mathbb{K}}$ denote the absolute Galois group of $\mathbb{K}$. A key ingredient of Wiles' celebrated proof of Fermat's Last Theorem is the following modularity lifting theorem:

Theorem 3.1.1 (Taylor-Wiles, R. Taylor and Wiles [1995]). *Let $p > 2$, $\mathcal{O}$ be the ring of integers in some finite extension of $\mathbb{Q}_p$, and let $\Phi : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathcal{O})$ be a Galois representation such that*

(i) *Φ "looks like" coming from a modular form;*

(ii) *The associated p-modular representation $\overline{\Phi} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is modular;*

(iii) *$\overline{\Phi}(G_{\mathbb{Q}(\sqrt{(-1)^{(p-1)/2}p})})$ is* **big**.

Then Φ is modular.

Here, *bigness* means just *irreducibility*.

In 2010, Harris, Shepherd-Barron, and R. Taylor [2010] proved the Sato-Tate conjecture for any non-CM elliptic curve over $\mathbb{Q}$ with non-integral j-invariant. A key role in this important result is played by an automorphy lifting theorem, due to Clozel, Harris, and

R. Taylor [2008], that generalizes the Taylor-Wiles Theorem 3.1.1 to GL_n. Now, *bigness* means irreducibility plus some more conditions, including a condition on the existence of a special element with a special *multiplicity-one* eigenvalue. The Clozel–Harris–Taylor theorem was generalized further by Thorne [2012], removing the multiplicity-one condition, and thus replacing *bigness* by *adequacy*:

Definition 3.1.2 (Thorne Thorne [2012] and Guralnick, Herzig, and Tiep [2015]**).** Let $\mathbb{F}$ be a field of characteristic p. A finite irreducible subgroup $G \leq \mathrm{GL}(V) = \mathrm{GL}_n(\mathbb{F})$ is called *adequate*, if

(A1) $H^1(G, \mathbb{F}) = 0$;
(A2) $H^1(G, \mathrm{End}(V)/\mathbb{F}) = 0$;
(A3) $\mathrm{End}(V)$ is linearly spanned by the p'-elements $g \in G$.

Which irreducible subgroups of $\mathrm{GL}(V)$ are adequate? Extending Guralnick, Herzig, R. Taylor, and Thorne [2012], and using results on Problem 1.4 as well as Blau and Zhang [1993], the following adequacy theorem has recently been proved:

Theorem 3.1.3. *Guralnick, Herzig, and Tiep [2015] Let $G < \mathrm{GL}(V)$ be a finite irreducible subgroup and let $\mathbf{O}^{p'}(G)$ denote the subgroup of G generated by all p-elements $x \in G$. Suppose that the $\mathbf{O}^{p'}(G)$-module V contains an irreducible submodule of dimension $< p$. Then, aside from a few explicitly described examples, G is adequate.*

This result has been extended in Guralnick, Herzig, and Tiep [2017] to include finite linear groups in dimension p. As a by-product, answers to a question of Serre concerning complete reducibility of subgroups in classical groups of low dimension, and a question of Mazur concerning $\dim \mathrm{Ext}^1(V, V)$ and $\dim \mathrm{Ext}^1(V, V^*)$ (which is of interest in deformation theory), have been obtained.

3.2 The α-invariant and Thompson's conjecture. Let $V = \mathbb{C}^n$ and let $G < GL(V)$ be a finite group. Then G acts on the dual space V^*, and a nonzero element $f \in \mathrm{Sym}^k(V^*)$ is said to be an *invariant*, respectively a *semi-invariant*, of degree k for G if G fixes f, respectively if G fixes the 1-dimensional space $\langle f \rangle_{\mathbb{C}}$. Let

$$d(G) := \min\{k \in \mathbb{N} | G \text{ has a semi-invariant of degree } k\}.$$

In 1981, Thompson proved the following theorem:

Theorem 3.2.1. *Thompson [1981] Let $n \in \mathbb{N}$ be any integer and $G < GL_n(\mathbb{C})$ be any finite subgroup. Then $d(G) \leq 4n^2$.*

It turns out that this result also has interesting implications in algebraic geometry, in particular, in regard to the α-invariant $\alpha_G(\mathbb{P}^{n-1})$ when $G < GL(V)$ acts on the projective space $\mathbb{P}V = \mathbb{P}^{n-1}$.

The *α-invariant* $\alpha_G(X)$ for a compact group G of automorphisms of a Kähler manifold X was introduced by Tian in 1987 Tian [1987], Tian and Yau [1987]. This invariant is of importance in differential geometry and algebraic geometry. As shown by Demailly and Kollár Demailly and Kollár [2001], in the case X is a Fano variety the α-invariant coincides with the *log-canonical threshold*

$$\mathrm{lct}(X,G) = \sup\left\{\lambda \in \mathbb{Q} \;\middle|\; \begin{array}{c}\text{the log pair } (X,\lambda D) \text{ has log-canonical singularities}\\ \text{for every } G\text{-invariant effective } \mathbb{Q}\text{-divisor } D \sim_{\mathbb{Q}} -K_X\end{array}\right\}.$$

An important example of Fano varieties is the projective space $\mathbb{P}V = \mathbb{P}^{n-1}$, where $V = \mathbb{C}^n$. Consider the natural action of any finite subgroup $G < GL(V)$ on $\mathbb{P}^{n-1}$. Then

$$\alpha_G(\mathbb{P}V) \leq \frac{d(G)}{\dim(V)},$$

see Cheltsov and Shramov [2011, §1], and so Theorem 3.2.1 implies

Theorem 3.2.2. *Thompson [1981] Let $n \in \mathbb{N}$ be any integer and $G < GL_n(\mathbb{C})$ be any finite subgroup. Then $\alpha_G(\mathbb{P}^{n-1}) \leq 4n$.*

In the same paper Thompson [ibid.], Thompson raised (the first part of) the following conjecture:

Conjecture 3.1 (Thompson). *There is a positive constant C such that for any $n \in \mathbb{N}$ and for any finite subgroup $G < GL_n(\mathbb{C})$, the following statements hold:*

(i) *$d(G) \leq Cn$; and*

(ii) *$\alpha_G(\mathbb{P}^{n-1}) \leq C$.*

This conjecture has recently been proved in Tiep [2016], relying on Aschbacher's theorem Aschbacher [1984] and results on Problem 1.4:

Theorem 3.2.3. *Tiep [2016] Thompson's Conjecture 3.1 is true, with $C = 1184036$.*

This implies

Corollary 3.2.4. *Let $G \leq GL(V)$ be a finite group. Then G has a nonzero polynomial invariant, of degree at most* $\min(1184036 \cdot \dim(V) \cdot \exp(G/G'), |G|)$.

3.3 Word maps on simple groups. The classical Waring problem, solved in 1909 by Hilbert, asks if given any $k \geq 1$, there is a (smallest) $g(k)$ such that every positive integer is a sum of at most $g(k)$ k^{th} powers. Recently, there has been considerable interest in non-commutative versions of the Waring problem, particularly for finite simple groups. Here, one asks if given any $k \geq 1$, there exists a (smallest) $f(k)$ such that every element in any finite non-abelian simple group G is a product of $f(k)$ k^{th} powers, provided that $\exp(G) \nmid k$. It was shown by Martinez and Zelmanov [1996] and independently by Saxl and Wilson [1997] that $f(k)$ exists (but implicitly).

More generally, given a *word*, i.e. an element $w(x_1, \ldots, x_d)$ of the free group $F(x_1, \ldots, x_d)$, and a group G, one considers the *word map* $w : G^d \to G$ and defines

$$w(G) := \{w(g_1, \ldots, g_d) \mid g_i \in G\}$$

to be the image of the word map w on G^d. Then the *non-commutative Waring problem* (for simple groups) can be formulated as follows

Problem 3.2. *Is there any integer c (possibly depending on w) such that*

$$w(G)^c := \{y_1 y_2 \ldots y_c \mid y_i \in w(G)\}$$

equals G, for any finite non-abelian simple group G with $w(G) \neq 1$?

Such smallest $c = c(w)$ is called the *width* of w. Of particular interest is the case where $w = w(x, y) = xyx^{-1}y^{-1}$, where the assertion $c(w) = 1$ is known as the *Ore conjecture* (1951), which is now a theorem:

Theorem 3.3.1. *Liebeck, O'Brien, Shalev, and Tiep [2010] Every element in any finite non-abelian simple group is a commutator.*

A particular motivation for Problem 3.2 comes from the celebrated Nikolov–Segal proof of the Serre conjecture on finitely generated profinite groups. The existence of $c(w)$ was first established in Liebeck and Shalev [2001] (again implicitly). Note that, in general, the width of w on simple groups can grow unbounded: as shown in Kassabov and Nikolov [2013] and Guralnick and Tiep [2015], for any $k \in \mathbb{N}$, there is a word w and a simple group S such that $w(S) \neq 1$ but $w(S)^N \neq S$. So in Problem 3.2 it is natural to bound the width of w on *sufficiently large* simple groups. In this asymptotic setting, a breakthrough was achieved by Shalev [2009], where he proved that *for any $w \neq 1$, $w(S)^3 = S$ for all sufficiently large simple groups S*. Building on Shalev [ibid.] and Larsen and Shalev [2008, 2009], the best solution for Problem 3.2 was achieved in Larsen, Shalev, and Tiep [2011]:

Theorem 3.3.2. *Larsen, Shalev, and Tiep [ibid.] The following statements hold.*

(i) *For any word $w \neq 1$, there exists a constant N_w depending on w, such that for all finite non-abelian simple groups S of order greater than N_w we have $w(S)^2 = S$.*

(ii) *For any two words $w_1, w_2 \neq 1$, there exists a constant N_{w_1,w_2} depending on w_1 and w_2 such that for all finite non-abelian simple groups S of order greater than N_{w_1,w_2} we have $w_1(S)w_2(S) = S$.*

As regards the Waring problem for quasisimple groups, the best solution has also been achieved:

Theorem 3.3.3. *Larsen, Shalev, and Tiep [2013] and Guralnick and Tiep [2015] The following statements hold.*

(i) *For any $w_1, w_2, w_3 \neq 1$, there exists a constant N_{w_1,w_2,w_3} depending on w_1, w_2, and w_3 such that for all finite quasisimple groups G of order greater than N_{w_1,w_2,w_3} we have $w_1(G)w_2(G)w_3(G) = G$.*

(ii) *For any $w_1, w_2 \neq 1$, there exists a constant N_{w_1,w_2} depending on w_1 and w_2 such that for all finite quasisimple groups G of order greater than N_{w_1,w_2} we have $w_1(G)w_2(G) \supseteq G \smallsetminus \mathbf{Z}(G)$.*

The aforementioned results on the Waring problem are mostly asymptotic and non-effective. Recently, effective versions of the main results of Martinez and Zelmanov [1996] and Saxl and Wilson [1997], as well as of Theorem 3.3.2 for power word maps, have been obtained:

Theorem 3.3.4. *Guralnick and Tiep [2015]*

(i) *Let $k \geq l \geq 1$. If S is any finite simple group of order $\geq k^{8k^2}$, then every $g \in S$ can be written as $x^k \cdot y^l$ for some $x, y \in S$.*

(ii) *Let $k \geq 1$ and let S be any finite simple group such that $\exp(S) \nmid k$. Then any element of S is a product of at most $80k\sqrt{2\log_2 k} + 56$ k^{th} powers in S.*

Also for power word maps, the following result has been proved, which generalizes classical theorems of Burnside and Feit–Thompson:

Theorem 3.3.5. *Guralnick, Liebeck, O'Brien, Shalev, and Tiep [n.d.]*

(i) *Let p, q be primes, let a, b be non-negative integers, and let $N = p^a q^b$. The word map $(x, y) \mapsto x^N y^N$ is surjective on all finite non-abelian simple groups.*

(ii) *Let N be an odd positive integer. The word map $(x, y, z) \mapsto x^N y^N z^N$ is surjective on all finite quasisimple groups.*

See also Guralnick, Liebeck, O'Brien, Shalev, and Tiep [ibid., Theorems 3–5] for results concerning power word maps $x \mapsto x^N$ for a general composite integer N.

3.4 Random walks, probabilistic generation, and representation varieties of Fuchsian groups. Let G be a finite group with a generating set S. Then the corresponding Cayley graph $\Gamma = \Gamma(G, S)$ has G as its vertex set and $\{(g, gs) \mid g \in G, s \in S\}$ as edge set. A *random walk* on Γ starts from 1, and at each step moves from a vertex g to gs chosen according to some probability distribution P on S. For any $t \in \mathbb{N}$, let $P^t(x)$ denote the probability of reaching $x \in G$ after t steps. A basic question is how fast P^t is converging to the uniform distribution U on G (i.e. $U(x) = 1/|G|$ for all $x \in G$), in the l^1-norm:

$$||P^t - U|| = \sum_{x \in G} |P^t(x) - U(x)|.$$

One then defines the *mixing time* $T = T(G, S)$ as

$$T = \min\{t \in \mathbb{N} \mid ||P^t - U|| < 1/e\}.$$

The study of random walks is pioneered by the influential work of Diaconis. A classical random walk was studied in Diaconis and Shahshahani [1981], where we want to shuffle a deck of n cards and at each shuffle we swap cards i and j with i, j chosen uniformly at random from $\{1, 2, \ldots, n\}$, that is, with $G = \mathsf{S}_n$ and $S = \{1, (ij) \mid 1 \leq i \neq n\}$. It was shown in Diaconis and Shahshahani [ibid.] that $T \approx (n \log n)/2$ for this card shuffle.

We will consider the case where $S = g^G = \{zgz^{-1} \mid z \in G\}$ and the distribution P is uniform: $P(s) = 1/|S|$ for all $s \in S$. In this case, the "Upper Bound Lemma" of Diaconis and Shahshahani [ibid.] states:

Lemma 3.4.1. *Suppose G is generated by $S = g^G$ and P is uniform on S. Then for any $t \in \mathbb{N}$,*

$$||P^t - U||^2 \leq \sum_{1_G \neq \chi \in \mathrm{Irr}(G)} \left(\frac{|\chi(g)|}{\chi(1)}\right)^{2t} \chi(1)^2.$$

Another tool allowing us to apply character-theoretic methods to these questions is provided by the following analogue of the Riemann zeta function

$$\zeta^G(s) = \sum_{\chi \in \mathrm{Irr}(G)} \chi(1)^{-s},$$

first considered by Witten [1991]. Now, sharp bounds on $\zeta^G(s)$ and exponential character bounds on $|\chi(g)|$ lead to strong results on mixing time $T(G, g^G)$, see e.g. Larsen and Shalev [2008] for the case of S_n.

Theorem 3.4.2. *Guralnick, Larsen, and Tiep [n.d.] Let $G = \mathrm{SL}_n^\epsilon(q)$ with $\epsilon = \pm$ and $n \geq 10$. Suppose that $g \in G$ is such that $|\mathbf{C}_{\mathrm{GL}_n^\epsilon(q)}(g)| \leq q^{n^2/12}$. Then $T(G, g^G) \leq 10$ for q sufficiently large.*

Proof. According to Theorem 2.5.4(iv), $|\chi(g)| \leq \chi(1)^{8/9}$. It then follows by Lemma 3.4.1 that

$$||P^t - U||^2 \leq \sum_{1_G \neq \chi \in \mathrm{Irr}(G)} \left(\frac{|\chi(g)|}{\chi(1)}\right)^{2t} \chi(1)^2 \leq \zeta^G(2t/9 - 2) - 1.$$

Taking $t \geq 10$, we have $2t/9 - 2 \geq 2/9 > 2/n$, whence $\lim_{q\to\infty} \zeta^G(2t/9 - 2) = 1$ by Liebeck and Shalev [2005a, Theorem 1.1], and the result follows. □

Similarly, the character bounds in Theorem 2.5.1 imply the following results on mixing time of random walks on quasisimple groups:

Theorem 3.4.3. *Bezrukavnikov, Liebeck, Shalev, and Tiep [n.d.] Suppose $\mathcal{G}$ is a simple algebraic group in good characteristic, and $G = G(q) = \mathcal{G}^F$ is a finite quasisimple group over $\mathbb{F}_q$. Let $g \in G$ be such that $\mathbf{C}_G(g) \leq L$, where $L = \mathcal{L}^F$ for a split Levi subgroup $\mathcal{L}$ of $\mathcal{G}$. Let $r = \mathrm{rank}(\mathcal{G})$ and let $h = (\dim \mathcal{G})/r - 1$ be the Coxeter number of $\mathcal{G}$.*

(i) *Suppose $\mathcal{G}$ is of classical type. Then the mixing time*

$$T(G, g^G) \leq \min\left\{r + 2, \left\lceil (2 + \frac{2}{h}) \cdot \frac{\dim \mathcal{G}}{\dim \mathcal{G} - \dim \mathcal{L}} \right\rceil\right\}$$

for large q.

(ii) *If $\mathcal{G}$ is of exceptional type, then $T(G, g^G) \leq 3$ for large q.*

Using Theorem 2.5.3, we obtain the following result covering all elements of $\mathrm{SL}_n(q)$:

Corollary 3.4.4. *Bezrukavnikov, Liebeck, Shalev, and Tiep [ibid.] Let $G = \mathrm{SL}_n(q)$ with $n \geq 5$ and let g be an arbitrary non-central element of G. Then the mixing time $T(G, g^G) \leq 2n + 3$ for large q.*

Next we discuss some applications to the study of representation varieties of Fuchsian groups, and also to probabilistic generation of finite simple groups. Recall that *Fuchsian groups* are finitely generated non-elementary discrete groups of isometries of the hyperbolic plane. Fuchsian groups, which include free groups, the modular group $\mathrm{PSL}_2(\mathbb{Z})$, surface groups, the Hurwitz group and hyperbolic triangle groups, play an important role in geometry, analysis, and algebra. Representation varieties of Fuchsian groups provide a convenient framework to generalize various results on random generation of finite simple groups (e.g. by two random elements, or by elements of orders 2 and 3). An old conjecture of G. Higman (now a theorem thanks to work of Conder, Everitt, and Liebeck and Shalev [2005b]) states that every Fuchsian group surjects onto all large enough alternating groups. Extending this, the following conjecture was raised by Liebeck and Shalev in Liebeck and Shalev [ibid.]:

Conjecture 3.3. *For any Fuchsian group Γ there is an integer $f(\Gamma)$, such that if G is a finite simple classical group of rank at least $f(\Gamma)$, then the probability $P_G(\Gamma)$ that a randomly chosen homomorphism from Γ to G is onto tends to 1 as $|G| \to \infty$.*

This conjecture was proved in Liebeck and Shalev [ibid.] for oriented Fuchsian groups of genus at least 2 and non-oriented Fuchsian groups of genus at least 3. Our new results on character bounds have allowed us to establish Conjecture 3.3 in various cases that had resisted all attacks so far.

Let Γ be a co-compact Fuchsian group of genus g having d elliptic generators of orders $m_1, \dots, m_d$ (all at least 2). Thus if Γ is oriented, it has a presentation of the following form:

$$\langle a_1, b_1, \dots, a_g, b_g,\ x_1, \dots, x_d \mid x_1^{m_1} = \dots = x_d^{m_d} = 1,\ x_1 \cdots x_d \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle,$$

and if Γ is non-oriented it has a presentation

$$\langle a_1, \dots, a_g,\ x_1, \dots, x_d \mid x_1^{m_1} = \dots = x_d^{m_d} = 1,\ x_1 \cdots x_d\, a_1^2 \cdots a_g^2 = 1 \rangle.$$

The *measure* of Γ is defined to be

$$\mu = \mu(\Gamma) := vg - 2 + \sum_{i=1}^{d} \left(1 - \frac{1}{m_i}\right) > 0,$$

where $v = 2$ if Γ is oriented and $v = 1$ otherwise. Let

$$N(\Gamma) := \max\left(\frac{2 + \sum \frac{1}{m_i}}{\mu},\ \frac{d+16}{4(\mu - 2)},\ m_1, \dots, m_d\right) + 1.$$

Theorem 3.4.5. *Liebeck, Shalev, and Tiep [n.d.] Let $\mathbb{K} = \overline{\mathbb{K}}$ be a field of characteristic not dividing $m_1 \cdots m_d$.*

(i) *If $\mu > 2$ and $n \geq N(\Gamma)$, then*

$$\dim \operatorname{Hom}(\Gamma, \mathrm{GL}_n(\mathbb{K})) = n^2(1 + \mu) - c,$$

where $-1 \leq c \leq \mu + 1 + \sum_{i=1}^{d} m_i$.

(ii) *Assume $\mu > \nu := \max\left(2,\ 1 + \sum \frac{1}{m_i}\right)$, and define*

$$\mathfrak{Q} := \bigcup_{primes\ p} \{q : q = p^a \equiv 1 (\operatorname{mod} m_i)\ \forall i\}.$$

Then for $n \geq \nu N(\Gamma) + 2\sum m_i$, *we have*

$$\lim_{q\to\infty,\, q\in\mathbb{Q}} P_\Gamma(\mathrm{SL}_n(q)) = 1.$$

(iii) *Let* $G(q)$ *denote a simple group of exceptional Lie type over* $\mathbb{F}_q$, *and suppose that* $\gcd(m_1\cdots m_d, 30) = 1$. *Then*

$$\lim_{q\to\infty,\, q\in\mathbb{Q},\, \gcd(q,30)=1} P_\Gamma(G(q)) = 1.$$

Theorem 3.4.5(iii) implies, for instance, that exceptional groups of Lie type $G(q)$, with $q \in \mathbb{Q}$ sufficiently large and $\gcd(m_1m_2m_3q, 30) = 1$, are images of the triangle group

$$T_{m_1,m_2,m_3} = \langle x_1, x_2, x_3 \mid x_1^{m_1} = x_2^{m_2} = x_3^{m_3} = x_1x_2x_3 = 1\rangle.$$

Results of this flavor on triangle generation were obtained by completely different methods in Larsen, Lubotzky, and Marion [2014a,b]. Further results concerning other finite groups of Lie type are also obtained in Liebeck, Shalev, and Tiep [n.d.].

References

Jonathan L. Alperin (1976). “The main problem of block theory”. In: *Proceedings of the Conference on Finite Groups (Univ. Utah, Park City, Utah, 1975)*. Academic Press, New York, pp. 341–356. MR: 0404417 (cit. on p. 244).

– (1987). “Weights for finite groups”. In: *The Arcata Conference on Representations of Finite Groups (Arcata, Calif., 1986)*. Vol. 47. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 369–379. MR: 933373 (cit. on p. 244).

Michael Aschbacher (1984). “On the maximal subgroups of the finite classical groups”. *Invent. Math.* 76.3, pp. 469–514. MR: 746539 (cit. on p. 256).

Vikraman Balaji and János Kollár (2008). “Holonomy groups of stable vector bundles”. *Publ. Res. Inst. Math. Sci.* 44.2, pp. 183–211. MR: 2426347 (cit. on p. 254).

Thomas R. Berger and Reinhard Knörr (1988). “On Brauer’s height 0 conjecture”. *Nagoya Math. J.* 109, pp. 109–116. MR: 931954 (cit. on p. 246).

Roman Bezrukavnikov, Martin W. Liebeck, Aner Shalev, and Pham Huu Tiep (n.d.). “Character bounds for finite groups of Lie type” (cit. on pp. 251, 252, 260).

Harvey I. Blau and Ji Ping Zhang (1993). “Linear groups of small degree over fields of finite characteristic”. *J. Algebra* 159.2, pp. 358–386. MR: 1231219 (cit. on p. 255).

Cédric Bonnafé and Raphaël Rouquier (2003). “Catégories dérivées et variétés de Deligne-Lusztig”. *Publ. Math. Inst. Hautes Études Sci.* 97, pp. 1–59. MR: 2010739 (cit. on p. 249).

Richard Brauer (1956). “Number theoretical investigations on groups of finite order”. In: *Proceedings of the international symposium on algebraic number theory, Tokyo and Nikko, 1955*. Science Council of Japan, Tokyo, pp. 55–62. MR: 0083490 (cit. on p. 244).

– (1963). “Representations of finite groups”. In: *Lectures on Modern Mathematics, Vol. I*. Wiley, New York, pp. 133–175. MR: 0178056 (cit. on p. 245).

Michel Broué and Jean Michel (1989). “Blocs et séries de Lusztig dans un groupe réductif fini”. *J. Reine Angew. Math.* 395, pp. 56–67. MR: 983059 (cit. on p. 249).

Jonathan Brundan and Alexander S. Kleshchev (2000). “Lower bounds for degrees of irreducible Brauer characters of finite general linear groups”. *J. Algebra* 223.2, pp. 615–629. MR: 1735167 (cit. on p. 249).

Ivan Cheltsov and Constantin Shramov (2011). “On exceptional quotient singularities”. *Geom. Topol.* 15.4, pp. 1843–1882. MR: 2860982 (cit. on p. 256).

Laurent Clozel, Michael Harris, and Richard Taylor (2008). “Automorphy for some l-adic lifts of automorphic mod l Galois representations”. *Publ. Math. Inst. Hautes Études Sci.* 108. With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras, pp. 1–181. MR: 2470687 (cit. on p. 254).

Everett C. Dade (1994). “Counting characters in blocks. II”. *J. Reine Angew. Math.* 448, pp. 97–190. MR: 1266748 (cit. on p. 247).

Jean-Pierre Demailly and János Kollár (2001). “Semi-continuity of complex singularity exponents and Kähler-Einstein metrics on Fano orbifolds”. *Ann. Sci. École Norm. Sup. (4)* 34.4, pp. 525–556. MR: 1852009 (cit. on p. 256).

Persi Diaconis and Mehrdad Shahshahani (1981). “Generating a random permutation with random transpositions”. *Z. Wahrsch. Verw. Gebiete* 57.2, pp. 159–179. MR: 626813 (cit. on p. 259).

Sergey Fomin and Nathan Lulov (1995). “On the number of rim hook tableaux”. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 223.Teor. Predstav. Din. Sistemy, Kombin. i Algoritm. Metody. I, pp. 219–226, 340. MR: 1374321 (cit. on pp. 243, 252).

Jason Fulman and Robert M. Guralnick (2012). “Bounds on the number and sizes of conjugacy classes in finite Chevalley groups with applications to derangements”. *Trans. Amer. Math. Soc.* 364.6, pp. 3023–3070. MR: 2888238 (cit. on p. 250).

Eugenio Giannelli, Alexander S. Kleshchev, Gabriel Navarro, and Pham Huu Tiep (2017). “Restriction of odd degree characters and natural correspondences”. *Int. Math. Res. Not. IMRN* 20, pp. 6089–6118. MR: 3712192 (cit. on p. 246).

Eugenio Giannelli, J. Tent, and Pham Huu Tiep (2018). “Irreducible characters of $3'$-degree of finite symmetric, general linear and unitary groups”. *J. Pure Appl. Algebra* 222 (8), pp. 2199–2228 (cit. on p. 246).

David Gluck (1993). "Character value estimates for nonsemisimple elements". *J. Algebra* 155.1, pp. 221–237. MR: 1206632 (cit. on p. 251).

– (1997). "Characters and random walks on finite classical groups". *Adv. Math.* 129.1, pp. 46–72. MR: 1458412 (cit. on p. 251).

Daniel Gorenstein, Richard Lyons, and Ronald Solomon (1994). *The classification of the finite simple groups*. Vol. 40. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. xiv+165. MR: 1303592 (cit. on p. 242).

Robert M. Guralnick, Florian Herzig, Richard Taylor, and Jack Thorne (2012). "Adequate subgroups". *J. Inst. Math. Jussieu* 11, pp. 907–920 (cit. on p. 255).

Robert M. Guralnick, Florian Herzig, and Pham Huu Tiep (2015). "Adequate groups of low degree". *Algebra Number Theory* 9.1, pp. 77–147. MR: 3317762 (cit. on p. 255).

– (2017). "Adequate subgroups and indecomposable modules". *J. Eur. Math. Soc. (JEMS)* 19.4, pp. 1231–1291. MR: 3626555 (cit. on p. 255).

Robert M. Guralnick, Mchael Larsen, and Pham Huu Tiep (n.d.). "Character levels and character bounds" (cit. on pp. 252–254, 259).

Robert M. Guralnick, Michael Larsen, and Pham Huu Tiep (2012). "Representation growth in positive characteristic and conjugacy classes of maximal subgroups". *Duke Math. J.* 161.1, pp. 107–137. MR: 2872555 (cit. on p. 248).

Robert M. Guralnick, Martin W. Liebeck, Eamonn A. O'Brien, Aner Shalev, and Pham Huu Tiep (n.d.). "Surjective word maps and Burnside's $p^a q^b$ theorem". To appear in Invent. Math. (2018) (cit. on p. 258).

Robert M. Guralnick, Kay Magaard, Jan Saxl, and Pham Huu Tiep (2002). "Cross characteristic representations of symplectic and unitary groups". *J. Algebra* 257.2, pp. 291–347. MR: 1947325 (cit. on p. 249).

Robert M. Guralnick, Gabriel Navarro, and Pham Huu Tiep (2016). "Finite groups with odd Sylow normalizers". *Proc. Amer. Math. Soc.* 144.12, pp. 5129–5139. MR: 3556259 (cit. on p. 246).

Robert M. Guralnick and Pham Huu Tiep (1999). "Low-dimensional representations of special linear groups in cross characteristics". *Proc. London Math. Soc. (3)* 78.1, pp. 116–138. MR: 1658160 (cit. on p. 249).

– (2004). "Cross characteristic representations of even characteristic symplectic groups". *Trans. Amer. Math. Soc.* 356.12, pp. 4969–5023. MR: 2084408 (cit. on p. 249).

– (2005). "Decompositions of small tensor powers and Larsen's conjecture". *Represent. Theory* 9, pp. 138–208. MR: 2123127 (cit. on p. 254).

– (2008). "Symmetric powers and a problem of Kollár and Larsen". *Invent. Math.* 174.3, pp. 505–554. MR: 2453600 (cit. on p. 254).

– (2012). "A problem of Kollár and Larsen on finite linear groups and crepant resolutions". *J. Eur. Math. Soc. (JEMS)* 14.3, pp. 605–657. MR: 2911879 (cit. on p. 254).

– (2015). "Effective results on the Waring problem for finite simple groups". *Amer. J. Math.* 137.5, pp. 1401–1430. MR: 3405871 (cit. on pp. 257, 258).

Michael Harris, Nick Shepherd-Barron, and Richard Taylor (2010). "A family of Calabi-Yau varieties and potential automorphy". *Ann. of Math. (2)* 171.2, pp. 779–813. MR: 2630056 (cit. on p. 254).

Gerhard Hiss and Gunter Malle (2001). "Low-dimensional representations of special unitary groups". *J. Algebra* 236.2, pp. 745–767. MR: 1813499 (cit. on p. 249).

I. Martin Isaacs (1973). "Characters of solvable and symplectic groups". *Amer. J. Math.* 95, pp. 594–635. MR: 0332945 (cit. on p. 246).

I. Martin Isaacs, Gunter Malle, and Gabriel Navarro (2007). "A reduction theorem for the McKay conjecture". *Invent. Math.* 170.1, pp. 33–101. MR: 2336079 (cit. on p. 245).

I. Martin Isaacs, Gabriel Navarro, Jørn B. Olsson, and Pham Huu Tiep (2017). "Character restrictions and multiplicities in symmetric groups". *J. Algebra* 478, pp. 271–282. MR: 3621673 (cit. on p. 246).

Gordon D. James (1983). "On the minimal dimensions of irreducible representations of symmetric groups". *Math. Proc. Cambridge Philos. Soc.* 94.3, pp. 417–424. MR: 720791 (cit. on p. 247).

Christoph Jansen (2005). "The minimal degrees of faithful representations of the sporadic simple groups and their covering groups". *LMS J. Comput. Math.* 8, pp. 122–144. MR: 2153793 (cit. on p. 247).

Martin Kassabov and Nikolai V. Nikolov (2013). "Words with few values in finite simple groups". *Q. J. Math.* 64.4, pp. 1161–1166. MR: 3151609 (cit. on p. 257).

Nicholas M. Katz (2004). "Larsen's alternative, moments, and the monodromy of Lefschetz pencils". In: *Contributions to automorphic forms, geometry, and number theory*. Johns Hopkins Univ. Press, Baltimore, MD, pp. 521–560. MR: 2058618 (cit. on p. 254).

– (2005). *Moments, monodromy, and perversity: a Diophantine perspective*. Vol. 159. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. viii+475. MR: 2183396 (cit. on p. 254).

Radha Kessar and Gunter Malle (2013). "Quasi-isolated blocks and Brauer's height zero conjecture". *Ann. of Math. (2)* 178.1, pp. 321–384. MR: 3043583 (cit. on p. 246).

– (2017). "Brauer's height zero conjecture for quasi-simple groups". *J. Algebra* 475, pp. 43–60. MR: 3612463 (cit. on pp. 246, 247).

Alexander S. Kleshchev and Pham Huu Tiep (n.d.). "Irreducible restrictions of representations of symmetric groups in small characteristics". In preparation (cit. on p. 248).

– (2004). "On restrictions of modular spin representations of symmetric and alternating groups". *Trans. Amer. Math. Soc.* 356.5, pp. 1971–1999. MR: 2031049 (cit. on p. 248).

– (2012). "Small-dimensional projective representations of symmetric and alternating groups". *Algebra Number Theory* 6.8, pp. 1773–1816. MR: 3033527 (cit. on p. 248).

János Kollár and Michael Larsen (2009). "Quotients of Calabi-Yau varieties". In: *Algebra, arithmetic, and geometry: in honor of Yu. I. Manin. Vol. II*. Vol. 270. Progr. Math. Birkhäuser Boston, Inc., Boston, MA, pp. 179–211. MR: 2641190 (cit. on p. 254).

Michael Larsen, Alexander Lubotzky, and Claude Marion (2014a). "Deformation theory and finite simple quotients of triangle groups I". *J. Eur. Math. Soc. (JEMS)* 16.7, pp. 1349–1375. MR: 3254328 (cit. on p. 262).

– (2014b). "Deformation theory and finite simple quotients of triangle groups II". *Groups Geom. Dyn.* 8.3, pp. 811–836. MR: 3267525 (cit. on p. 262).

Michael Larsen, Gunter Malle, and Pham Huu Tiep (2013). "The largest irreducible representations of simple groups". *Proc. Lond. Math. Soc. (3)* 106.1, pp. 65–96. MR: 3020739 (cit. on p. 250).

Michael Larsen and Aner Shalev (2008). "Characters of symmetric groups: sharp bounds and applications". *Invent. Math.* 174.3, pp. 645–687. MR: 2453603 (cit. on pp. 243, 251, 257, 259).

– (2009). "Word maps and Waring type problems". *J. Amer. Math. Soc.* 22.2, pp. 437–466. MR: 2476780 (cit. on p. 257).

Michael Larsen, Aner Shalev, and Pham Huu Tiep (2011). "The Waring problem for finite simple groups". *Ann. of Math. (2)* 174.3, pp. 1885–1950. MR: 2846493 (cit. on pp. 251, 257).

– (2013). "Waring problem for finite quasisimple groups". *Int. Math. Res. Not. IMRN* 10, pp. 2323–2348. MR: 3061941 (cit. on p. 258).

Martin W. Liebeck, Eamonn A. O'Brien, Aner Shalev, and Pham Huu Tiep (2010). "The Ore conjecture". *J. Eur. Math. Soc. (JEMS)* 12.4, pp. 939–1008. MR: 2654085 (cit. on p. 257).

Martin W. Liebeck and Aner Shalev (2001). "Diameters of finite simple groups: sharp bounds and applications". *Ann. of Math. (2)* 154.2, pp. 383–406. MR: 1865975 (cit. on p. 257).

– (2005a). "Character degrees and random walks in finite groups of Lie type". *Proc. London Math. Soc. (3)* 90.1, pp. 61–86. MR: 2107038 (cit. on p. 260).

– (2005b). "Fuchsian groups, finite simple groups and representation varieties". *Invent. Math.* 159.2, pp. 317–367. MR: 2116277 (cit. on pp. 260, 261).

Martin W. Liebeck, Aner Shalev, and Pham Huu Tiep (n.d.). "Character ratios, representation varieties and random generation of finite groups of Lie type" (cit. on pp. 261, 262).

Benjamin F. Logan and Lawrence A. Shepp (1977). "A variational problem for random Young tableaux". *Advances in Math.* 26.2, pp. 206–222. MR: 1417317 (cit. on p. 242).

George Lusztig (1984). *Characters of reductive groups over a finite field*. Vol. 107. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xxi+384. MR: 742472 (cit. on pp. 242, 249).

– (1988). “On the representations of reductive groups with disconnected centre”. *Astérisque* 168. Orbites unipotentes et représentations, I, pp. 10, 157–166. MR: 1021495 (cit. on pp. 242, 249).

Gunter Malle and Britta Späth (2016). “Characters of odd degree”. *Ann. of Math. (2)* 184.3, pp. 869–908. MR: 3549625 (cit. on p. 245).

Consuelo Martinez and Efim Zelmanov (1996). “Products of powers in finite simple groups”. *Israel J. Math.* 96.part B, pp. 469–479. MR: 1433702 (cit. on pp. 257, 258).

John McKay (1972). “Irreducible representations of odd degree”. *J. Algebra* 20, pp. 416–418. MR: 0286904 (cit. on p. 244).

Masafumi Murai (2012). “On Brauer’s height zero conjecture”. *Proc. Japan Acad. Ser. A Math. Sci.* 88.3, pp. 38–40. MR: 2908621 (cit. on p. 247).

Gabriel Navarro (2003). “A new character correspondence in groups of odd order”. *J. Algebra* 268.1, pp. 8–21. MR: 2004477 (cit. on p. 246).

– (2004). “The McKay conjecture and Galois automorphisms”. *Ann. of Math. (2)* 160.3, pp. 1129–1140. MR: 2144975 (cit. on p. 245).

Gabriel Navarro and Britta Späth (2014). “On Brauer’s height zero conjecture”. *J. Eur. Math. Soc. (JEMS)* 16.4, pp. 695–747. MR: 3191974 (cit. on p. 247).

Gabriel Navarro and Pham Huu Tiep (2011). “A reduction theorem for the Alperin weight conjecture”. *Invent. Math.* 184.3, pp. 529–565. MR: 2800694 (cit. on p. 246).

– (2012). “Brauer’s height zero conjecture for the 2-blocks of maximal defect”. *J. Reine Angew. Math.* 669, pp. 225–247. MR: 2980457 (cit. on p. 247).

– (2013). “Characters of relative p'-degree over normal subgroups”. *Ann. of Math. (2)* 178.3, pp. 1135–1171. MR: 3092477 (cit. on p. 247).

Gabriel Navarro, Pham Huu Tiep, and Alexandre Turull (2007). “p-rational characters and self-normalizing Sylow p-subgroups”. *Represent. Theory* 11, pp. 84–94. MR: 2306612 (cit. on p. 245).

Gabriel Navarro, Pham Huu Tiep, and Carolina Vallejo (2014). “McKay natural correspondences on characters”. *Algebra Number Theory* 8.8, pp. 1839–1856. MR: 3285617 (cit. on p. 246).

Alexander A. Premet (1987). “Weights of infinitesimally irreducible representations of Chevalley groups over a field of prime characteristic”. *Mat. Sb. (N.S.)* 133(175).2, pp. 167–183, 271. MR: 905003 (cit. on p. 249).

Lluis Puig (2011). “On the reduction of Alperin’s conjecture to the quasi-simple groups”. *J. Algebra* 328, pp. 372–398. MR: 2745572 (cit. on p. 246).

Jan Saxl and John S. Wilson (1997). “A note on powers in simple groups”. *Math. Proc. Cambridge Philos. Soc.* 122.1, pp. 91–94. MR: 1443588 (cit. on pp. 257, 258).

Amanda A. Schaeffer Fry (2016). “Odd-degree characters and self-normalizing Sylow 2-subgroups: a reduction to simple groups”. *Comm. Algebra* 44.5, pp. 1882–1904. MR: 3490653 (cit. on p. 246).

Amanda A. Schaeffer Fry and Jay Taylor (2018). “On self-normalising Sylow 2-subgroups in type A”. *J. Lie Theory* 28.1, pp. 139–168. MR: 3694099 (cit. on p. 246).

Gary M. Seitz (1990). “Cross-characteristic embeddings of finite groups of Lie type”. *Proc. London Math. Soc. (3)* 60.1, pp. 166–200. MR: 1023808 (cit. on p. 250).

Aner Shalev (2009). “Word maps, conjugacy classes, and a noncommutative Waring-type theorem”. *Ann. of Math. (2)* 170.3, pp. 1383–1416. MR: 2600876 (cit. on p. 257).

Britta Späth (2013a). “A reduction theorem for the Alperin-McKay conjecture”. *J. Reine Angew. Math.* 680, pp. 153–189. MR: 3100954 (cit. on pp. 245, 247).

– (2013b). “A reduction theorem for the blockwise Alperin weight conjecture”. *J. Group Theory* 16.2, pp. 159–220. MR: 3031870 (cit. on p. 246).

Richard Taylor and Andrew Wiles (1995). “Ring-theoretic properties of certain Hecke algebras”. *Ann. of Math. (2)* 141.3, pp. 553–572. MR: 1333036 (cit. on p. 254).

John G. Thompson (1981). “Invariants of finite groups”. *J. Algebra* 69.1, pp. 143–145. MR: 613863 (cit. on pp. 255, 256).

Jack Thorne (2012). “On the automorphy of l-adic Galois representations with small residual image”. *J. Inst. Math. Jussieu* 11.4. With an appendix by Robert Guralnick, Florian Herzig, Richard Taylor and Jack Thorne, pp. 855–920. MR: 2979825 (cit. on p. 255).

Gang Tian (1987). “On Kähler-Einstein metrics on certain Kähler manifolds with $C_1(M) > 0$”. *Invent. Math.* 89.2, pp. 225–246. MR: 894378 (cit. on p. 256).

Gang Tian and Shing-Tung Yau (1987). “Kähler-Einstein metrics on complex surfaces with $C_1 > 0$”. *Comm. Math. Phys.* 112.1, pp. 175–203. MR: 904143 (cit. on p. 256).

Pham Huu Tiep (2003). “Low dimensional representations of finite quasisimple groups”. In: *Groups, combinatorics & geometry (Durham, 2001)*. World Sci. Publ., River Edge, NJ, pp. 277–294. MR: 1994973 (cit. on p. 249).

– (2006). “Finite groups admitting Grassmannian 4-designs”. *J. Algebra* 306.1, pp. 227–243. MR: 2271581 (cit. on p. 249).

– (2016). “The α-invariant and Thompson’s conjecture”. *Forum Math. Pi* 4, e5, 28. MR: 3519762 (cit. on p. 256).

Pham Huu Tiep and Alexander E. Zalesskii (1996). “Minimal characters of the finite classical groups”. *Comm. Algebra* 24.6, pp. 2093–2167. MR: 1386030 (cit. on p. 249).

Anatolli M. Vershik and Sergei V. Kerov (1985). “Asymptotic behavior of the maximum and generic dimensions of irreducible representations of the symmetric group”. *Funktsional. Anal. i Prilozhen.* 19.1, pp. 25–36, 96. MR: 783703 (cit. on p. 242).

Edward Witten (1991). “On quantum gauge theories in two dimensions”. *Comm. Math. Phys.* 141.1, pp. 153–209. MR: 1133264 (cit. on p. 259).

Received 2017-11-16.

Pham Huu Tiep

Department of Mathematics
Rutgers University
Piscataway, NJ 08854-8019
USA
tiep@math.rutgers.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (267–294)

p-ADIC VARIATION OF AUTOMORPHIC SHEAVES

Fabrizio Andreatta, Adrian Iovita and Vincent Pilloni

Abstract

We review the construction of analytic families of Siegel modular cuspforms based on the notion of overconvergent modular forms of p-adic weight. We then present recent developments on the following subjects: the halo conjecture, the construction of p-adic L-functions, and the modularity of irregular motives.

1 Introduction

We start by fixing a number field F, a prime integer $p > 0$ and an integer $n \geq 1$ and by denoting G_F and respectively $\mathbb{A}_F$ the absolute Galois group and the ring of adeles of F. We fix an isomorphism $\mathbb{C} \simeq \overline{\mathbb{Q}}_p$. One of the most mysterious conjectures in number theory, known as the *Langlands, Clozel, Fontaine-Mazur Conjecture* is the statement of the existence of a bijection respecting L-functions between the following families of isomorphic classes of representations:

$\mathrm{Rep}^{\mathrm{geom}}_{p,F} :=$ {Irreducible representations $\rho\colon G_F \to \mathrm{GL}_n(\bar{\mathbb{Q}}_p)$ which are continuous, ramified only at a finite number of places and de Rham at the places dividing p}

and

$\mathrm{Aut}^{\mathrm{alg}}_F :=$ {Cuspidal algebraic automorphic representations π of the group $\mathrm{GL}_n(\mathbb{A}_F)$}.

The algebraicity condition of an automorphic form π is the condition that the infinitesimal character (also called the weight) of the π_v is algebraic for all infinite places v of F (Clozel [1990], Buzzard and Gee [2011]). It is known that there are only countably many isomorphism classes of algebraic automorphic representations (Harish-Chandra [1968], Thm. 1). On the other hand the objects in $\mathrm{Rep}^{\mathrm{geom}}_{p,F}$ should arise from the cohomology of proper smooth varieties over F (Fontaine and Mazur [1995], Conj. 1). The bijection we are seeking should therefore be a bijection of (conjecturally) countable sets.

Having written this let us remark that in fact each of these countable sets can be embedded in certain analytic varieties over $\mathbb{Q}_p$. To be more precise, for $\mathrm{Rep}^{\mathrm{geom}}_{p,F}$, one can

MSC2010: primary 11G18; secondary 11G40.

relax the condition of the representations being de Rham at places dividing p and obtain a moduli space of more general, non-geometric p-adic Galois representations enjoying reasonable finiteness properties (Mazur [1989]).

Concerning $\mathrm{Aut}_F^{\mathrm{alg}}$, one would also like to see them as a subset of a larger set of p-adic automorphic forms. For the moment there are several good definitions of a p-adic automorphic form depending on the way one endows the space $\mathrm{Aut}_F^{\mathrm{alg}}$ with a p-adic topology.

The general method to study elements of $\mathrm{Aut}_F^{\mathrm{alg}}$ is to realize them (when this is possible!) in the Betti cohomology of a locally symmetric space or in the coherent cohomology of a Shimura variety (Harris [1990]). These cohomology groups naturally carry structures of finite dimensional $\mathbb{Q}$-vector spaces and these structures can be used to equip the automorphic forms with a p-adic topology. Once a cohomological realization has been chosen, one can start to vary the levels and the weights of the algebraic automorphic representations. Miraculously, we find in many cases that the countable set of systems of eigenvalues associated to the cuspidal algebraic automorphic forms is not isolated in the space of p-adic automorphic forms and that its closure acquires the structure of an analytic space.

The point of view of p-adically deforming the elements of $\mathrm{Rep}_{p,F}^{\mathrm{geom}}$ and $\mathrm{Aut}_F^{\mathrm{alg}}$ proves to be very fruitful as it is sometimes easier to study and work with these analytic varieties than to work with the geometric Galois representations and algebraic automorphic forms individually.

In this note we begin by explaining one possible approach for the construction of the p-adic analytic spaces (also called eigenvarieties) attached to p-adic automorphic forms based on the coherent cohomology realization in a Shimura variety of PEL type[1]. We shall actually limit ourselves to the Siegel moduli spaces of polarized abelian varieties with level structure. We explain how to vary p-adically the weight of the automorphic forms. One has as a guiding principle that in order to be able to deform automorphic forms one needs to allow them, seen as global sections of certain automorphic sheaves, to have essential singularities at non-ordinary points. Restricting the automorphic vector bundles to the complement of these non-ordinary points has the advantage that they (the automorphic vector bundles) acquire extra structures arising from the universal p-divisible group via the Hodge-Tate period map. Our main task is to define overconvergent modular forms of any p-adic weight. This is a refinement of the definition of p-adic modular forms of Serre, Katz [1973] and Hida [1986] and an interpolation of the notion of overconvergent modular forms of integral weight considered by Dwork, Coleman and Mazur [1998]... This material has already appeared in print (for example in Andreatta, Iovita, and Pilloni [2015]).

[1]This choice is made to the expense of ignoring interesting automorphic forms because the condition of admitting a cohomological realization in the coherent cohomology of a Shimura variety is restrictive. There are other approaches based on the Betti realization but for the sake of brevity we will not discuss them here.

Next we present three recent developments in this area: the halo conjecture, the construction of triple product p-adic L functions in the finite slope case, and the modularity of certain irregular motives.

2 Vector bundles with marked sections

In this section we review some constructions which applied to various contexts provide examples of interpolation of automorphic sheaves in the subsequent sections. We have tried to isolate the key representation theoretic ideas outlining a general method to get interpolations that might be useful in other situations.

2.1 Some classical representation theory. We start by recalling the construction of the irreducible representations of the group GL_g. Let $\mathrm{B} \subset \mathrm{GL}_g$ be the upper triangular Borel and let T be the usual diagonal torus. Let $X(\mathrm{T})$ be the character group of T. This group is isomorphic to $\mathbb{Z}^g$ via the map sending $(k_1, \cdots, k_g)$ to the character $\mathrm{diag}(t_1, \cdots, t_g) \mapsto \prod_{i=1}^{g} t_i^{k_i}$. We let $X(\mathrm{T})^+$ be the cone of dominant weights given by the condition $k_1 \geq k_2 \cdots \geq k_g$.

If $k \in X(\mathrm{T})^+$, we define the algebraic induction $V^k = \{f : \mathrm{GL}_g \to \mathbb{A}^1,\ f(gb) = k(b)f(g),\ \forall (g, b) \in \mathrm{GL}_g \times \mathrm{B}\}$ where k has been extended to a character of B via the projection $\mathrm{B} \to \mathrm{T}$. The group GL_g acts on this space via $f \mapsto f(g \cdot -)$.

2.2 p-Adic representation theoretic variations. Next we explain how to interpolate the weights k and the spaces V^k, for $k \in X(\mathrm{T})^+$. We let $\Lambda = \mathbb{Z}_p[\![T(\mathbb{Z}_p)]\!]$ be the Iwasawa algebra of the torus. The universal continuous character of $\mathrm{T}(\mathbb{Z}_p)$ is the tautological character:

$$k^{\mathrm{un}} \colon \mathrm{T}(\mathbb{Z}_p) \longrightarrow \Lambda^\times.$$

We can consider the formal spectrum $\mathfrak{W} = \mathrm{Spf}\,\Lambda$ and denote by $\mathcal{W}$ its rigid analytic fiber over $\mathrm{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$. This is a finite union of open unit polydiscs of dimension g. Given a complete Huber pair (B, B^+) over $(\mathbb{Q}_p, \mathbb{Z}_p)$, the morphisms $\mathrm{Spa}(B, B^+) \to \mathcal{W}$ correspond to $\mathrm{Hom}_{\mathrm{cont}}(\mathrm{T}(\mathbb{Z}_p), B^\times)$. In particular $\mathcal{W}(\mathbb{Q}_p)$ contains the algebraic weights $X(\mathrm{T})$. Observe that $X(\mathrm{T})$ is totally disconnected in $\mathcal{W}$ but Zariski dense and that $\mathcal{W}$ has only a finite number of connected components.

The character k^{un} interpolates the algebraic characters $X(\mathrm{T})$. We now explain how to interpolate the representations $\{V^k\}_{k \in X(\mathrm{T})^+}$ over $\mathcal{W}$ (Stevens [2000]). We switch to the analytic setting and let now GL_g, T, B denote the analytifications over $\mathrm{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$ of the respective group schemes. Let $\mathrm{Iw} \subset \mathrm{GL}_g(\mathbb{Z}_p)$ be the Iwahori subgroup of matrices which are upper triangular modulo p. For any number $w \in \mathbb{Q}_{>0} \cup \{\infty\}$, we denote by $\mathrm{Iw}_w \subset \mathrm{GL}_g$ the adic analytic subgroup of GL_g of integral matrices which are congruent modulo p^w to an element of Iw and we let $\mathrm{B}_w = \mathrm{B} \cap \mathrm{Iw}_w$ and $\mathrm{T}_w = \mathrm{T} \cap \mathrm{Iw}_w$. Let $\mathcal{U} \hookrightarrow \mathcal{W}$ be an open subspace. Let w be such that the universal character extends to a pairing $k^{\mathrm{un}}_{\mathcal{U}} \colon \mathrm{T}_w \times \mathcal{U} \to \mathbb{G}_m$. In this case we say that k^{un} is w-analytic over $\mathcal{U}$. Remark that if $\mathcal{U}$ is quasi-compact, then $k^{\mathrm{un}}_{\mathcal{U}}$ is always w-analytic for some $w \in \mathbb{Q}_{>0}$. We may define a representation of the group Iw_w as the analytic induction

$$V_w^{k^{\mathrm{un}}_{\mathcal{U}}} = \{ f : \mathrm{Iw}_w \times \mathcal{U} \to \mathbb{A}^1 \times \mathcal{U},\ \forall (i,b) \in \mathrm{Iw}_w \times \mathrm{B}_w \ \ f(ib) = k^{\mathrm{un}}_{\mathcal{U}}(b) f(i) \}.$$

If $\mathcal{U} = \mathrm{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$ and $k^{\mathrm{un}}_{\mathcal{U}} = k \in X(\mathrm{T})^+$ (k is algebraic and therefore w-analytic for all w) we have an inclusion $V^k \hookrightarrow V_w^{k^{\mathrm{un}}_{\mathcal{U}}}$. Observe that unless $g = 1$, the space $V_w^{k^{\mathrm{un}}_{\mathcal{U}}}$ is an infinite dimensional Banach space and the inclusion is not an isomorphism. This should not be a surprise as for $g \geq 2$ the dimensions of the spaces $\{V^k\}_{k \in X(\mathrm{T})^+}$ vary and the only possibility to interpolate them is to embed them in larger spaces (infinite dimensional) which can then be interpolated. It is moreover possible to characterize V^k inside $V_w^{k^{\mathrm{un}}_{\mathcal{U}}}$ by using some differential operators (analytic BGG resolution, Jones [2011]).

2.3 Relative constructions. The classical case. We use the notation of Section 2.1. Let X be a scheme, let $\mathcal{E}$ be a locally free sheaf of rank g over X and denote by $\mathcal{E}^\vee = \underline{\mathrm{Hom}}(\mathcal{E}, \mathcal{O}_X)$ the dual sheaf. We associate to any dominant weight $k \in X(\mathrm{T})^+$ a locally free sheaf $\mathcal{E}^k$ over X as follows. Consider the GL_g-torsor $f : \mathcal{T}(\mathcal{E}) \to X$ associated to $\mathcal{E}$, namely $\mathcal{T}(\mathcal{E}) := \mathrm{Isom}(\mathcal{O}_X^g, \mathcal{E}^\vee)$. Define $\mathcal{E}^k = f_* \mathcal{O}_{\mathcal{T}(\mathcal{E})}[k]$, the functions on $\mathcal{T}(\mathcal{E})$ transforming via k under the action of B. One gets a finite, locally free $\mathcal{O}_X$-module which locally on X is isomorphic to the space V^k introduced in Section 2.1.

2.4 Relative constructions. *p*-Adic variations. We now assume that $\mathcal{X}$ is an analytic adic space over $\mathrm{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$. Let $\mathcal{E}$ be a locally free sheaf of rank g over $\mathcal{X}$ and let $\mathcal{E}^+$ be an integral structure, namely a subsheaf of finite and locally free $\mathcal{O}^+_{\mathcal{X}}$-modules of rank g such that $\mathcal{E} = \mathcal{E}^+ \otimes_{\mathcal{O}^+_{\mathcal{X}}} \mathcal{O}_{\mathcal{X}}$. Let $w \in \mathbb{Q}_{>0} \cup \{\infty\}$. We now provide a formalism which leads to the construction of families of sheaves interpolating the sheaves $\{\mathcal{E}^k\}_{k \in X(\mathrm{T})^+}$ on $\mathcal{X}$ and which, locally on $\mathcal{X}$, are isomorphic to the spaces V_w^k of Section 2.2. The new essential ingrediants are the "marked sections" $s_1, \dots, s_g \in \mathrm{H}^0(\mathcal{X}, \mathcal{E}^+/p^w\mathcal{E}^+)$ with the property that the induced map $(\mathcal{O}^+_{\mathcal{X}}/p^w\mathcal{O}^+_{\mathcal{X}})^g \to \mathcal{E}^+/p^w\mathcal{E}^+$ is bijective.

Define $\mathcal{T}_w(\mathcal{E}^+, \{s_1, \dots, s_g\})$ as the functor that associates to any adic space $t : \mathcal{Z} \to \mathcal{X}$ the set of sections $(\rho_1, \cdots, \rho_g) \in \mathrm{H}^0\big(\mathcal{Z}, t^*(\mathcal{E}^+)^\vee\big)$ such that $(\langle t^*(s_i), \rho_j \rangle)_{1 \leq i,j \leq g} \in \mathrm{Iw}$ mod p^w. Here $t^\star(\mathcal{E}^+)$ is the sheaf $t^{-1}(\mathcal{E}^+) \otimes_{t^{-1}\mathcal{O}^+_{\mathcal{X}}} \mathcal{O}^+_{\mathcal{Z}}$ and $t^\star(\mathcal{E}^+)^\vee$ is its $\mathcal{O}^+_{\mathcal{Z}}$-dual. One proves easily that $\mathcal{T}_w(\mathcal{E}^+, \{s_1, \dots, s_g\})$ is representable by an adic space and is a Iw_w-torsor. We now assume that there is a map $\mathcal{X} \to \mathcal{W}$ and that the character $k_{\mathcal{X}}$ pulled back from the universal character on $\mathcal{W}$ is w-analytic. Under this assumption we define the sheaf

$$\mathcal{E}_w^k := f_* \mathcal{O}_{\mathcal{T}_w(\mathcal{E}^+, \{s_1, \ldots, s_g\})}[k_{\mathfrak{X}}].$$

This sheaf is a relative version of the construction of $V_w^{k_{\mathfrak{X}}}$ given in Section 2.2.

We now describe a slight variant of this construction where we only assume that we have a partial set of sections. In this situation it is still possible to realize a partial interpolation. Let $1 \leq r \leq g$. We define the subgroup $\mathrm{Iw}_{w,r}$ of GL_g of integral matrices of the form

$$\begin{pmatrix} A & D \\ B & C \end{pmatrix},$$

where $A \in \mathrm{GL}_r$ and $A \mod p^w$ is upper triangular with entries in $\mathbb{Z}_p/p^w$, $D \in \mathrm{M}_{r,g-r}$, $C \in \mathrm{GL}_{g-r}$, $B \in \mathrm{M}_{g-r,r}$ and $B = 0 \mod p^w$. We denote by $\mathrm{T}_{w,r} = \mathrm{T} \cap \mathrm{Iw}_{w,r}$, $\mathrm{B}_{w,r} = \mathrm{B} \cap \mathrm{Iw}_{w,r}$ and $\mathrm{T}_{w,r} = \mathrm{T} \cap \mathrm{Iw}_{w,r}$. We assume that we have sections $s_1, \cdots, s_r \in \mathrm{H}^0(\mathfrak{X}, \mathcal{E}^+/p^w \mathcal{E}^+)$ such that the induced map $(\mathcal{O}_{\mathfrak{X}}^+/p^w \mathcal{O}_{\mathfrak{X}}^+)^r \to \mathcal{E}^+/p^w \mathcal{E}^+$ is injective with locally free cokernel of rank $g - r$. We define $\mathcal{T}_w(\mathcal{E}^+, \{s_1, \ldots, s_r\})$ as the functor that associates to any adic space $t : \mathcal{Z} \to \mathfrak{X}$ the set of basis $(\rho_1, \cdots, \rho_g) \in \mathrm{H}^0(\mathcal{Z}, t^\star(\mathcal{E}^+)^\vee)$ such that:

- $(\langle t^\star(s_i), \rho_j \rangle)_{1 \leq i,j \leq r} \in \mathrm{GL}_r(\mathbb{Z}_p) \mod p^w$ and is upper triangular modulo p^w,
- $\langle t^\star(s_i), \rho_j \rangle = 0 \mod p^w$ for all $1 \leq i \leq r$ and $g - r + 1 \leq j \leq g$.

It is clear that $\mathcal{T}_w(\mathcal{E}^+, \{s_1, \ldots, s_r\})$ is an $\mathrm{Iw}_{w,r}$-torsor. We now assume that the character $k_{\mathfrak{X}}$ extends to a character of $\mathrm{T}_{w,r}$ and we denote by $\mathcal{E}_w^k := f_* \mathcal{O}_{\mathcal{T}_w(\mathcal{E}^+, \{s_1, \ldots, s_r\})}[k_{\mathfrak{X}}]$.

We remark that the relative constructions in Sections 2.2 and 2.4 could have been made exactly in the same way by working with an invertible ideal $I \subset \mathcal{O}_{\mathfrak{X}}^+$ such that $I \cap \mathbb{Z}_p = p^w \mathbb{Z}_p$, instead of with $p^w \mathcal{O}_{\mathfrak{X}}^+$.

3 Variations in the Siegel case

Let GSp_{2g} be the group of similitudes of $(\mathbb{Z}^{2g}, \langle , \rangle)$ where $\langle , \rangle$ is the alternating form given by $\langle e_i, e_{2g-i+1} \rangle = 1$ if $1 \leq i \leq g$ and $\langle e_i, e_j \rangle = 0$ if $i + j \neq 2g + 1$. Let $K \subset \mathrm{GSp}_{2g}(\mathbb{A}_f)$ be a neat compact open subgroup, where $\mathbb{A}_f$ denotes the ring of finite adels of the rationals. Let $Y_K \to \mathrm{Spec}\ \mathbb{Q}$ be the Siegel moduli space of polarized abelian varieties A of dimension g and level structure K. Its complex analytification $(Y_K \times \mathrm{Spec}\ \mathbb{C})^{an}$ is the locally symmetric space $\mathrm{GSp}_{2g}(\mathbb{Q}) \backslash \big(\mathcal{H}_g \times \mathrm{GSp}_{2g}(\mathbb{A}_f)/K\big)$ where $\mathcal{H}_g = \{M \in \mathrm{M}_g(\mathbb{C}), M^t = M,\ \mathrm{Im}(M)$ is definite positive or negative$\}$ is the Siegel space i.e. the union of the Siegel upper and lower half-spaces.

3.1 The classical construction. To any g-uple $k = (k_1, \dots, k_g) \in \mathbb{Z}^g$ satisfying $k_1 \geq k_2 \geq \dots \geq k_g$, one attaches an automorphic locally free sheaf ω^k on Y_K using the construction of Section 2.3 for the sheaf ω_A of invariant differentials of the universal abelian scheme over Y_K. Let X_K be a toroidal compactification of Y_K (Faltings and Chai [1990]). The sheaf ω^k extends canonically to X_K. The global sections form the space of classical holomorphic Siegel modular forms of weight k and level K. This is a finite dimensional $\mathbb{Q}$-vector space. It carries an action of the Hecke algebra $\mathcal{C}_c^\infty(\mathrm{GSp}_{2g}(\mathbb{A}_f)//K, \mathbb{Z})$ of locally constant and compactly supported functions which are left and right K invariant on $\mathrm{GSp}_{2g}(\mathbb{A}_f)$.

After tensoring with $\mathbb{C}$, these Siegel modular forms can be described as holomorphic vector valued functions on $\mathcal{H}_g$ satisfying a transformation property with respect to a congruence subgroup of $\mathrm{GSp}_{2g}(\mathbb{Q})$. The cuspidal forms (those vanishing on $D = X_K \setminus Y_K$) define (via a usual lifting process) special vectors in the space of algebraic automorphic forms for the group $\mathrm{GSp}_{2g}(\mathbb{A})$. Here and elsewhere $\mathbb{A}$ denotes the ring of adels of $\mathbb{Q}$.

3.2 Interpolation. Let $p > 0$ be a prime integer. We now assume that $K = K^p K_p$ where $K^p \subset \mathrm{GSp}_{2g}(\mathbb{A}_f^{(p)})$ and $K_p = \mathrm{GSp}_{2g}(\mathbb{Z}_p)$. In this setting, Y_K and X_K admit canonical models over Spec $\mathbb{Z}_{(p)}$ and we denote by $\mathcal{Y}$ and respectively $\mathcal{X}$ the associated analytic spaces over $\mathrm{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$. Over $\mathcal{Y}$ there is a universal p-divisible group $A[p^\infty]$, which comes with a quasi-polarisation: $A[p^\infty] \simeq (A[p^\infty])^D$.

We review the method of Andreatta, Iovita, and Pilloni [2015] to construct a sheaf interpolating the classical automorphic sheaves ω^k. We shall work over $\mathcal{Y}$ for simplicity, but everything extends to $\mathcal{X}$. This construction relies on the Hodge-Tate period map

$$\mathrm{HT} : T_p(A) \to \omega_A$$

where $T_p(A)$ is the Tate-module of the p-divisible group $A[p^\infty]$, a pro-étale sheaf locally isomorphic to $\mathbb{Z}_p^{2g}$. Over the ordinary locus $\mathcal{Y}^{\mathrm{ord}}$ we have an étale-multiplicative extension $0 \to T_p(A)^m \to T_p(A) \to T_p(A)^{et} \to 0$ and the Hodge-Tate map factors through a map $T_p(A)^{et} \to \omega_A$ which induces an isomorphism of pro-étale sheaves $T_p(A)^{et} \otimes \mathcal{O}_{\mathcal{Y}^{\mathrm{ord}}} \to \omega_A|_{\mathcal{Y}^{\mathrm{ord}}}$. Thus the GL_g-torsor ω_A arises from a $\mathrm{GL}_g(\mathbb{Z}_p)$-torsor over the ordinary locus and this allows the interpolation of the sheaves ω^k over $\mathcal{Y}^{\mathrm{ord}}$. It is nevertheless important in order to have compact operators and for the construction of eigenvarieties to go beyond the ordinary locus.

Given an integer $r \geq 0$ we let $\mathcal{Y}_r \subset \mathcal{Y}$ be the open defined by the valuations x satisfying the inequality $|\widetilde{\mathrm{Ha}}^{p^{r+1}}|_x \geq |p|_x$ where $\widetilde{\mathrm{Ha}}$ is locally defined as a (any) lift of the Hasse invariant on the special fiber of Y_K. Each $\mathcal{Y}_r$ should be thought of as a tubular neighborhood of the ordinary locus $\mathcal{Y}^{\mathrm{ord}}$ in $\mathcal{Y}$, where $\mathcal{Y}^{\mathrm{ord}}$ is defined by the condition that $|\widetilde{\mathrm{Ha}}|_x \geq 1$. It follows from the theory of the canonical subgroup that the p^r-torsion of A

over $\mathcal{Y}_r$ contains a canonical subgroup $H_r \subset A[p^r]$ (see Fargues [2011]). Over $\mathcal{Y}^{\mathrm{ord}}$ it coincides with the multiplicative part of $A[p^r]$.

In order to apply the general machinery of Section 2.4 we need to exhibit a vector bundle with marked sections. Consider the finite étale cover of adic spaces $\mathcal{IG}_r \to \mathcal{Y}_r$ classifying trivializations $\psi : (\mathbb{Z}/p^r\mathbb{Z})^g \cong H_r^D$. Then $\mathcal{IG}_r$ carries several sheaves:

1) we have the sheaf H_r and its Cartier dual H_r^D;

2) we have a sheaf ω_A^+, resp. ω_A of $\mathcal{O}^+_{\mathcal{IG}_r}$-modules, resp. of $\mathcal{O}_{\mathcal{IG}_r}$-modules, which are locally free and finite of rank g. Over affinoids $\mathrm{Spa}(B, B^+) \subset \mathcal{IG}_r$ such that the pull-back of A extends to an abelian scheme $\widetilde{A}$ over B^+, the value of ω_A^+ and of ω_A are the module of invariant differentials of $\widetilde{A}$, resp. of A;

3) we have a sheaf $\omega^+_{H_r}$ of $\mathcal{O}^+_{\mathcal{IG}_r}$-modules and a morphism $\mathrm{HT}\colon H_r^D \to \omega^+_{H_r}$. Over affinoids $\mathrm{Spa}(B, B^+) \subset \mathcal{IG}_r$ such that the pull-back of H_r extends to a finite and flat group scheme $\widetilde{H}_r \hookrightarrow \widetilde{A}$ over B^+, the value of $\omega^+_{H_r}$ is the module of invariant differentials $\omega_{\widetilde{H}_r}$ and the map HT is the Hodge-Tate map.

Notice that we have a natural morphism $\omega_A^+ \to \omega^+_{H_r}$. With this we define a *modification* $\omega_A^\sharp \subset \omega_A^+$ as the inverse image in ω_A^+ of $\mathrm{HT}\big(H_r^D \otimes \mathcal{O}^+_{\mathcal{IG}_r}\big)$. One proves that this is a finite and locally free sheaf of $\mathcal{O}^+_{\mathcal{IG}_r}$-modules over $\mathcal{IG}_r$ of rank g and that for every rational number $0 < w \le r - \frac{1}{p(p-1)}$ the morphism HT defines an isomorphism of $\mathcal{O}^+_{\mathcal{IG}_r}/p^w$-modules

$$\mathrm{HT}_w\colon H_r^D \otimes \mathcal{O}^+_{\mathcal{IG}_r}/p^w \cong \omega_A^\sharp/p^w\omega_A^\sharp.$$

which is a good substitute of the comparison map we had over the ordinary locus.

Consider $\mathcal{E} := \omega_A$, $\mathcal{E}^+ := \omega_A^\sharp$ and the sections $s_1, \ldots, s_g$ of $\omega_A^\sharp/p^w\omega_A^\sharp$ provided by the images of the canonical basis of $(\mathbb{Z}/p^r\mathbb{Z})^g$ via $\mathrm{HT}_w \circ \psi$. Let $\mathcal{U} \subset \mathcal{W}$ be an open subset where the character $k^{un}_{\mathcal{U}}$ is w-analytic. Applying the construction explained in Section 2.4 we get the sheaves we are looking for

$$\omega^{k^{\mathrm{un}}}_{r,\mathcal{U}} := \pi_\star\big(\mathcal{O}_{\mathcal{T}_w(\omega_A^\sharp,\{s_1,\ldots,s_g\})\times\mathcal{U}}\big)[k^{\mathrm{un}}],$$

where $\pi : \mathcal{T}_w(\omega_A^\sharp, \{s_1, \ldots, s_g\}) \to \mathcal{IG}_r$ is the torsor of trivializations of $\omega_A^\sharp$ with marked sections $s_1, \ldots, s_g$.

Actually, denote by $\mathcal{Y}_{\mathrm{Iw},r} \to \mathcal{Y}_r$ be the covering parametrizing full flags of H_1^D. Then the sheaf $\omega^{k^{\mathrm{un}}}_{r,\mathcal{U}}$ descends canonically along the natural map $\mathcal{IG}_r \to \mathcal{Y}_{\mathrm{Iw},r}$. Moreover it extends without much difficulties to the toroidal compactification $\mathcal{X}_{\mathrm{Iw},r}$ of $\mathcal{Y}_{\mathrm{Iw},r}$.

3.2.1 A perfectoid digression. We'd like to explain the construction of the previous section when $g = 1$ in more elementary terms. Let N be an integer, $N \ge 3$ and $(N, p) = 1$.

Recall (see Katz [1973]) that a modular form f of weight $k \in \mathbb{Z}$ with level $\Gamma_1(N)$ over $\mathbb{Z}[1/N]$ can be viewed as a functorial rule mapping a triple $(E/R, P, \omega)$ (consisting of an elliptic curve $E \to \operatorname{Spec} R$ for a $\mathbb{Z}[1/N]$-algebra R, a point $P \in E[N]$ of order exactly N, and a nowhere vanishing differential form ω) to $f(E/R, P, \omega) \in R$ satisfying the additional transformation property: $f(E, P, \lambda.\omega) = \lambda^{-k} f(E, P, \omega)$ for any $\lambda \in R^\times$ and some growth condition at infinity.

Over $\mathbb{C}$, we can pull back f to a function on the Poincaré upper half plane by setting $F(\tau) = f(\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z}), \frac{1}{N}, \mathrm{d}z)$ for the coordinate z on $\mathbb{C}$. For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N)$, multiplication by $(c\tau+d)^{-1}$ on $\mathbb{C}$ identifies the triples $(\mathbb{C}/(\mathbb{Z}+\tau\mathbb{Z}), \frac{1}{N}, \mathrm{d}z)$ and $(\mathbb{C}/(\mathbb{Z}+\gamma.\tau\mathbb{Z}), \frac{1}{N}, (c\tau + d)\mathrm{d}z)$ and therefore F satisfies a descent condition with respect to the action of $\Gamma_1(N)$, namely $F(\gamma.\tau) = (c\tau + d)^k F(\tau)$.

We now express our definition of overconvergent modular forms of some $\mathbb{C}_p$-valued character $k : \mathbb{Z}_p^\times \to \mathbb{C}_p^\times$ in similar terms. Assume that k is w-analytic and choose a positive integer r such that $r-1 < w \le r - \frac{1}{(p-1)p}$ (this can be achieved at the expense of increasing w). Then an r-overconvergent modular form f of weight k is a rule associating to every quadruple (E, P, ψ, ω) an element $f(E, P, \psi, \omega) \in \mathbb{C}_p$, where E is an elliptic curve over $\mathbb{C}_p$ such that $|\widetilde{\mathrm{Ha}}^{p^{r+1}}(E)| \ge |p|$, P is a point of order N, ψ is point of order p^r of the dual canonical subgroup H_r^D, ω is an integral differential form on E such that $\omega \mod p^w = \mathrm{HT}_w(\Psi)$. Moreover we demand that f is "analytic", extends to the cusps, and satisfies the functional equation $f(E, P, \lambda\psi, \lambda\omega) = k^{-1}(\lambda) f(E, P, \psi, \omega)$ for all $\lambda \in \mathbb{Z}_p^\times(1 + p^w\mathbb{C}_p)$.[2]

Following Chojecki, Hansen, and Johansson [2017], one can describe an analogue of the passage from f to F in the p-adic world. Let $\mathfrak{X}(\infty) \to \mathfrak{X}$ be the prefectoid modular curve of level $\Gamma(p^\infty) \cap \Gamma_1(N)$ constructed by Scholze [2015] and $\mathcal{Y}(\infty)$ the complement of the boundary. Over $\mathcal{Y}(\infty)$ we have a universal trivialization $\psi_\infty \colon \mathbb{Z}_p^2 \cong T_p(E)$ and the Hodge-Tate map $\mathrm{HT}\colon T_p(E) \to \omega_E$ induces a period map:

$$\pi_{\mathrm{HT}} \colon \mathfrak{X}(\infty) \longrightarrow \mathbb{P}^1$$

which, over $\mathcal{Y}(\infty)$, is characterized by the fact that $\pi_{\mathrm{HT}}^\star\big(\mathcal{O}_{\mathbb{P}^1}(1)\big) = \omega_E$ and the pull-back of the two canonical sections s_0 and s_1 of $\mathcal{O}_{\mathbb{P}^1}(1)$ are the images via $\mathrm{HT} \circ \psi_\infty$ of the canonical basis e_0, e_1 of $\mathbb{Z}_p \oplus \mathbb{Z}_p$. For any $v \in \mathbb{Q}_{>0}$, let $\mathbb{P}^1_v$ be the open of $\mathbb{P}^1$ defined by the condition $|s_1| \le |p^v s_0|$.

Let $\mathfrak{X}(\infty)_v = \pi_{\mathrm{HT}}^{-1}(\mathbb{P}^1_v)$ and $\mathcal{Y}(\infty)_v = \mathcal{Y}(\infty) \cap \mathfrak{X}(\infty)_v$. For v large enough, $(E, P, \psi_\infty) \in \mathcal{Y}(\infty)_v$ has a canonical subgroup of level r which is generated by the image of $\psi_\infty(e_1)$ in

[2] In order to make sense of the functional equation it is necessary to restrict to differential forms which "arise" from the dual canonical subgroup

$E[p^r]$, and $\psi_\infty(e_0)$ maps to a generator $\overline{\psi_\infty(e_0)}$ of H_r^D. We can therefore pullback f to a function on $\mathcal{Y}(\infty)_v$ by setting $F(E, P, \psi_\infty) = f(E, P, \overline{\psi_\infty(e_0)}, s_0 = \mathrm{HT}(\psi_\infty(e_0)))$.

This identifies the space of overconvergent modular forms of weight k with a space of functions on the open $\mathfrak{X}(\infty)_v$ of $\mathfrak{X}(\infty)$. These functions satisfy a descent condition which reminds us of the descent condition on the upper half plane. Namely, let n be the smallest integer greater than v. We consider the subgroup $K_0(p^n) \subset \mathrm{GL}_2(\mathbb{Z}_p)$ of elements $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $c \in p^n\mathbb{Z}_p$. For any $\gamma \in K_0(p^n)$ as above, we find that $F(E, P, \psi_\infty \circ \gamma) = k^{-1}(a + b\frac{s_1}{s_0})F(E, P, \psi_\infty)$. [3]

3.3 Eigenvarieties. The sheaves $\omega_{r,\mathfrak{U}}^{k^{\mathrm{un}}}$ produce variations of Hecke eigensystems as follows. The global sections of $\omega_{r,\mathfrak{U}}^{k^{\mathrm{un}}}$ over $\mathfrak{X}_{\mathrm{Iw},r}$, vanishing at the boundary, form the Banach module of r-overconvergent, w-analytic cuspidal Siegel modular forms of weight parametrized by $\mathfrak{U}$. Passing to the limit over r and w we obtain the space of overconvergent, locally analytic cuspidal Siegel modular forms of weight parametrized by $\mathfrak{U}$. Let N be the product of primes different from p for which $K_\ell \neq \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$. This space carries an action of the commutative spherical Hecke algebra $\mathbb{T}^{Np} := \mathcal{C}_c^\infty(\mathrm{GSp}_{2g}(\mathbb{A}_f^{(Np)})//K^{Np}, \mathbb{Z})$. Let $\mathrm{Iw}_p \subset \mathrm{GSp}_{2g}(\mathbb{Z}_p)$ be the Iwahori parahoric of upper triangular matrices modulo p. At p, there is an action of the dilating Hecke algebra $\mathbb{U}_p := \mathbb{Z}[U_{p,1}, \cdots, U_{p,g}]$[4] $\subset \mathcal{C}_c^\infty(\mathrm{GSp}_{2g}(\mathbb{Q}_p)//\mathrm{Iw}_p, \mathbb{Z})$, and the operator $U = \prod_i U_{p,i}$ is compact.

Let f be a classical cuspidal eigenform of weight k and level $K^p\mathrm{Iw}_p$. We denote by $\Theta_f : \mathbb{T}^{Np} \otimes \mathbb{U}_p \to \bar{\mathbb{Q}}$ the associated character.[5] We have the following:

Theorem 3.1. *There is a rigid analytic space $\mathcal{E}$, called the eigenvariety of tame level K^p, equipped with a weight map $w : \mathcal{E} \to \mathcal{W}$ which is locally on the source and the target finite and torsion free and there is a universal Hecke character $\Theta : \mathbb{T}^{Np} \otimes \mathbb{U}_p \to \mathcal{O}_\mathcal{E}$ with dense image such that:*

- *Any classical cuspidal eigenform f of weight k and level $K^p\mathrm{Iw}_p$ provides a unique point x_f on $\mathcal{E}$ such that $\Theta|_{x_f} = \Theta_f$ and $w(x_f) = k$,*
- *Conversely, any point $x \in \mathcal{E}$ satisfying $w(x) = (k_1, \ldots, k_g) \in X(\mathrm{T})^+$ satisfying $v(\Theta|_x(U_{p,i})) < k_{g-i} - k_{g-i+1} + 1$ for $1 \leq i \leq g-1$ and $v(\Theta|_x(U_{p,g})) < k_g - \frac{g(g+1)}{2}$ arises from a cuspidal eigenform f of weight k and level $K^p\mathrm{Iw}_p$.*

[3] $F(E, P, \psi_\infty \circ \gamma) = f(E, P, \overline{\psi_\infty(ae_0)}, \mathrm{HT} \circ \psi_\infty(ae_0 + be_1)) = f(E, P, \overline{\psi_\infty(ae_0)}, (a + b\frac{s_1}{s_0})\mathrm{HT} \circ \psi_\infty(e_0)) = k^{-1}(a + b\frac{s_1}{s_0})F(E, P, \psi_\infty)$.

[4] If $i \in \{1, \cdots, g-1\}$, $U_{p,i}$ is the characteristic function of the double class $\mathrm{Iw}_p\mathrm{diag}(p^2 Id_i, pId_{2g-2i}, Id_i)\mathrm{Iw}_p$, $U_{p,g}$ is the characteristic function of the double class $\mathrm{Iw}_p\mathrm{diag}(pId_g, Id_g)\mathrm{Iw}_p$

[5] Since f has Iwahori level at p, then $\Theta_f(U_{p,i}) \neq 0$ and f is of finite slope.

Remark 3.3.1. 1) The case $g = 1$ of Theorem 3.1 was first proved by Coleman and Mazur in Coleman and Mazur [1998]. They used a different construction of the p-adic families of modular forms in which the Eisenstein family plays a crucial role and which could not be generalized for $g > 1$.

2) The cuspidality condition is crucial for the theorem for $g \geq 2$. We prove an acyclicity result for the cuspidal sheaves $\omega^{k^{\mathrm{un}}}_{r,\mathcal{U}}(-D)$ using that $\mathcal{X}_{\mathrm{Iw,r}}$ has affine image in the minimal compactification and showing that the relative cohomology of cuspidal sheaves between the toroidal and the minimal compactifications vanishes in degrees greater than 1. In particular the acyclicity allows us to prove that the degree zero cohomology of $\omega^{k^{\mathrm{un}}}_{r,\mathcal{U}}(-D)$ commutes with specializations in the weight space $\mathcal{U} \subset \mathcal{W}$.

3) We outlined the construction for Siegel modular varieties but the same method applies more generally for PEL type Shimura varieties having dense ordinary locus, see Andreatta, Iovita, and Pilloni [2016a] for the Hilbert case and Brasca [2016] for the general case.

4) Even for Shimura varieties with empty ordinary locus, one can proceed in a similar way. The ordinary locus is replaced by the so called μ-ordinary locus, introduced by T. Wedhorn, and the Hasse invariant is replaced by the μ-Hasse invariant, defined at various levels of generality by G. Boxer, W. Goldring-M.H. Nicole, V. Hernandez, J.S. Koskivirta-T. Wedhorn. The last ingredient one needs is a replacement for the canonical subgroup and the Hodge-Tate map. We refer to Kassaei [2013] and Brasca [2013] for the case of Shimura curves and to Hernandez [2016] for the more general case of PEL type Shimura varieties and for a thorough account of the problem.

5) The last point of the Theorem 3.1 is proven in Bijakowski, Pilloni, and Stroh [2016] (and already by Coleman and Kassaei for $g = 1$). It is a *classicity criterion* which roughly asserts that small slope overconvergent modular forms are classical. It is a crucial result in order to study eigenvarieties as it provides a dense set of points, the classical ones. In a certain sense, these classical points characterize uniquely the eigenvariety. In particular, it often happens that a given property known at the classical points can be inferred by continuity for the whole eigenvariety.

4 Variations at infinity

We now restrict to the case $g = 1$. In this case we have an eigencurve $\mathcal{E} \to \mathcal{Z} \to \mathcal{W}$ where $\mathcal{Z} \hookrightarrow \mathcal{W} \times \mathbb{G}_m$, the so called *spectral curve*, is the zero locus of the characteristic series $\mathcal{P}(X)$ of the U-operator acting on the space of overconvergent modular forms. The map $\mathcal{E} \to \mathcal{Z}$ is finite and both $\mathcal{E}$ and $\mathcal{Z}$ are equidimensional of dimension 1. Therefore the geometry of $\mathcal{E}$ can be understood, to some extent, by studying the apparently simpler space $\mathcal{Z}$.

4.1 The spectral halo. Recall that the weight space $\mathcal{W}$ is the rigid analytic fiber of $\operatorname{Spf} \mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]$. One can consider a slightly bigger space $\mathcal{W}^{\mathrm{an}}$ defined by the analytic points of the adic space $\operatorname{Spa}(\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!], \mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!])$. Recall that

$$\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!] \cong \mathbb{Z}_p\big[(\mathbb{Z}/p\mathbb{Z})^\times\big][\![T]\!]$$

where T is defined by imposing that the grouplike element $\exp(p)$ is equal to $T+1$. The complement of $\mathcal{W}$ in $\mathcal{W}^{\mathrm{an}}$ consists of finitely many points in characteristic p, corresponding to the T-adic valuations on $\mathbb{F}_p\big[(\mathbb{Z}/p\mathbb{Z})^\times\big](\!(T)\!)$ and $\mathcal{W}^{\mathrm{an}}$ is a compactification of $\mathcal{W}$, obtained by adding a point at the boundary of each rigid analytic open unit disc.

Coleman observed that the characteristic series $\mathcal{P}(X)$ of the U-operator on the eigencurve has coefficients in Λ and, hence one can consider the extended spectral curve $\mathcal{Z}^{an} \hookrightarrow \mathcal{W}^{an} \times \mathbb{G}_m = V(\mathcal{P})$. The fiber of $\mathcal{Z}^{an}$ over a boundary point

$$k\colon \operatorname{Spa}\big(\mathbb{F}_p(\!(T)\!), \mathbb{F}_p[\![T]\!]\big) \to \mathcal{W}^{an}$$

is the zero set of the specialization $\mathcal{P}_k(X)$ at k, over the non-archimedean field $\mathbb{F}_p(\!(T)\!)$.

In Andreatta, Iovita, and Pilloni [2018] we prove a conjecture of Coleman in which he stated the existence of a Banach space over $\mathbb{F}_p(\!(T)\!)$ and of a compact operator whose characteristic series is $\mathcal{P}_k(X)$. More precisely, we prove the following result. Let $\mathcal{X}_{\mathcal{W}^{\mathrm{an}}}$ be the analytic adic space defined by the pull back of the modular curve to $\mathcal{W}^{\mathrm{an}}$. Given $v \in \mathbb{Q}_{\geq 0}^\times$ we define $\mathcal{X}_{\mathcal{W}^{\mathrm{an}}}(v)$ to be the open consisting of the points x satisfying the condition $|\widetilde{\mathrm{Ha}}|_x \geq \sup\{|T^v|_x, |p^v|_x\}$.

Theorem 4.1. *For $v > 0$ small enough we have an invertible sheaf $\omega^{k^{\mathrm{un}}}$ over $\mathcal{X}_{\mathcal{W}^{\mathrm{an}}}(v)$, endowed with an action of the Hecke operators, that coincides with the construction in Section 3.2 over* $\operatorname{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$.

Moreover, given a boundary point k of $\mathcal{W}^{\mathrm{an}}$, the sections of the fiber of $\omega^{k^{\mathrm{un}}}$ at k form a Banach module over $\mathbb{F}_p(\!(T)\!)$ such that the characteristic series of the U-operator is Coleman's series $\mathcal{P}_k(X)$.

The sections of the characteristic p fiber of $\omega^{k^{\mathrm{un}}}$ are called *T-adic overconvergent modular forms* (of radius of convergence v). They are actually functions on certain overconvergent Igusa tower in characteristic p. Using the sheaf $\omega^{k^{\mathrm{un}}}$ over $\mathcal{X}_{\mathcal{W}^{\mathrm{an}}}(v)$ one manages to extend the Coleman-Mazur eigencurve to an eigencurve $\mathcal{E}^{\mathrm{an}}$ over the whole $\mathcal{Z}^{\mathrm{an}}$.

Each finite slope eigenform f in characteristic p defines a point on $\mathcal{E}^{\mathrm{an}}$ and we can associate to it a semi-simple two dimensional Galois representation

$$\rho_f\colon \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2\big(\overline{\mathbb{F}_p[\![T]\!]}\big)$$

unramified at the primes different form p and not dividing the tame level. Here $\overline{\mathbb{F}_p[\![T]\!]}$ denotes an algebraic closure of $\mathbb{F}_p[\![T]\!]$. If f is ordinary, ρ_f has already been constructed by Hida. For finite slopes we get new, mysterious objects in the realm of Galois representations that deserve further study and understanding. Here are some questions that we find interesting.

Given ρ_f as above, one can construct an equicharacteristic p, étale (φ, Γ)-module $\mathfrak{D}(\rho)$ over the Robba ring for the discretely valued field $\bar{\mathbb{F}}_p((t))$ (as in Berger and Colmez [2008]). Is $\mathfrak{D}(\rho)$ trianguline (i.e., extension of one dimensional (φ, Γ)-modules)?

Does this characterize the two dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ with values in finite extensions of $\bar{\mathbb{F}}_p((T))$ which arise from T-adic overconvergent modular eigenforms of finite slope?

4.2 The halo conjecture. We would now like to discuss the halo conjecture and some questions related to the global geometry of $\mathcal{E}^{an}$. Let $x\colon \mathrm{Spa}(K, \mathcal{O}_K) \to \mathcal{W}^{an}$ be a rank one point. The choice of a pseudo uniformizer ϖ allows to normalize the associated valuation $v\colon K \to \mathbb{R} \cup \{\infty\}$ by $v(\varpi) = 1$. In that case we write v_ϖ for v. There are in general two natural choices of pseudo-uniformizer in $\mathcal{O}_K$, namely p and T, except at the boundary when $p = 0$ and at the very center $T = 0$. One can attach to the characteristic series $\mathcal{P}_k(X) = \sum_{n \geq 0} a_n X^n$ and a choice of pseudo uniformizer ϖ a Newton polygon $NP_\varpi(\mathcal{P}_k)$ which is the convex enveloppe of the points $(n, v_\varpi(a_n)) \subset \mathbb{R}^2$. This is the graph of a piecewise linear function and the sequence of slopes of $NP_\varpi(\mathcal{P}_k)$ is giving the sequence of ϖ-adic valuations of finite slope eigenvalues of U.

Conjecture 1 (Coleman-Mazur-Buzzard-Kilford)**.** *Let $k \in \mathcal{W}^{an}$ be a boundary point. Then there exists a positive rational number r such that for all rank one points*

$$k'\colon \mathrm{Spa}\big(K, \mathcal{O}_K\big) \to \mathcal{W}^{an}$$

in a neighbourhood $\mathcal{U} = \{x,\ |p^r|_x \leq |T|_x\}$ of k we have $NP_T(\mathcal{P}_{k'}) = NP_T(\mathcal{P}_k)$. Moreover, the slopes in $NP_T(\mathcal{P}_k)$ form a finite union of arithmetic progressions.

Before discussing what is known about this conjecture, let us describe some of the consequences. The first implication is that $\mathcal{Z}^{an}|_{\mathcal{U}} = \coprod_s \mathcal{Z}^{an}(s)$ splits as a disjoint union of components according to the slopes s occurring in $NP_T(\mathcal{P}_k)$. Each component $\mathcal{Z}^{an}(s)$ is finite flat over $\mathcal{U}$. In certain numerical examples it actually maps isomorphically onto $\mathcal{U}$. In this case the complement of the points at infinity can be visualized as halos, explaining the name of the conjecture.

A second implication is that the p-adic slopes tend to zero as one approaches the boundary. In particular, T-adic overconvergent modular eigenforms of finite slope (for the U-operator) are limits of classical modular forms of arbitrary fixed weight $k \geq 2$ (of course of increasing level at p) by Coleman's classicity theorem. It is known that each irreducible component of $\mathcal{Z}^{an}$ has image in the weight space equal to the complement of a finite number of points. Therefore, any irreducible component of $\mathcal{Z}^{an}$ contains at least one irreducible component of some $\mathcal{Z}^{an}(s)$. Thus each irreducible component of $\mathcal{Z}^{an}$ contains infinitely many classical points of a given weight $k \geq 2$.

Another consequence of the fact that the p-adic slopes tend to zero approaching infinity is that an irreducible component of $\mathcal{Z}^{an}$ is finite over the weight space if and only if it is ordinary (i.e., the slope is 0) (Liu, Wan, and Xiao [2017], prop. 3.24).

The conjecture has not yet been proved for the whole eigencurve $\mathcal{E}^{an}$ but it is known for all the irreducible components that arise from the p-adic Jaquet-Langlands correspondence thanks to Liu, Wan, and Xiao [ibid.]. Independently, in Bergdall and Pollack [2016] it is proved that the constancy of the Newton Polygon implies the second part of the halo conjecture, namely that the slopes form a finite union of arithmetic progressions.

Motivated by the boundary behavior provided by the conjecture and by a conjecture of Buzzard's for classical weights, Bergdall and Pollack have elaborated a unifying conjecture in Bergdall and Pollack [2017], called *ghost conjecture*, predicting (under some extra assumptions) the slopes of overconvergent cuspforms over the whole weight space.

Finally let us remark that, even though we discussed only the elliptic case, eigenvarieties might be defined over the whole analytic adic weight space for more general Shimura varieties. We refer to Andreatta, Iovita, and Pilloni [2016b] for the Hilbert case and to Johansson and Newton [2016] for a Betti cohomology approach. In contrast with the elliptic case, where at infinity the weight space consists of a finite set of points, in the Hilbert case, for a totally real field of degree g, at infinity the weight space has components of dimension $g-1$.

5 p-Adic variation of de Rham automorphic sheaves

In this section we use the notations and results of Section 2 and Section 3 for $G = \mathrm{GL}_2/\mathbb{Q}$, i.e., for $g = 1$. Here we briefly present the constructions and results of Andreatta and Iovita [2017], using adic analytic spaces instead of formal schemes. The interested reader should consult Andreatta and Iovita [ibid.] for more details.

Before getting into technicalities let us briefly explain the problem we are faced with and explain how we chose to solve it. Let $p > 2$ be a prime integer, $N \geq 3$ an integer relatively prime to p, $\mathcal{X}$ the adic analytic projective modular curve of level $\Gamma_1(N)$ over $\mathrm{Spa}(\mathbb{Q}_p, \mathbb{Z}_p)$ and $\alpha\colon E \longrightarrow \mathcal{X}$ the generalized, universal elliptic curve. We denote by $\left(\mathrm{H}_E, \mathrm{Fil}_\bullet, \nabla\right)$ the data consisting of:

i) the relative de Rham cohomology sheaf of E over $\mathcal{X}$, i.e.

$$\mathrm{H}_E := \mathbb{R}^1\alpha_*\left(\Omega^\bullet_{E/\mathcal{X}}(\log(\alpha^{-1}(\mathrm{cusps})))\right),$$

ii) the Hodge filtration $\mathrm{Fil}_\bullet$ of H_E, i.e., $\mathrm{Fil}_0 := \omega_E = \alpha_*\big(\Omega^1_{E/\mathfrak{X}}(\log(\alpha^{-1}(\mathrm{cusps})))\big)$, $\mathrm{Fil}_i = 0$ for $i < 0$ and $\mathrm{Fil}_i = \mathrm{H}_E$ for $i \geq 1$.

iii) $\nabla\colon \mathrm{H}_E \longrightarrow \mathrm{H}_E \otimes_{\mathcal{O}_{\mathfrak{X}}} \Omega^1_{\mathfrak{X}/\mathbb{Q}_p}\big(\log(\mathrm{cusps})\big)$, the Gauss-Manin connection, an integrable connection satisfying Griffith's transversality property.

We now consider the following family of data indexed by the integers:

$$(*) \qquad \Big(\big(\mathrm{Sym}^n(\mathrm{H}_E), \mathrm{Fil}_{n,\bullet}, \nabla_n\big)\Big)_{n\in\mathbb{Z}},$$

where $\mathrm{Fil}_{n,\bullet}$ and ∇_n are the natural increasing filtrations and connections on the n-th symmetric powers of H_E.

Over the complex numbers one can use the Hodge decomposition of H_E in order to describe the global sections of $\omega_E^{k-r} \otimes \mathrm{Sym}^r(\mathrm{H}_E)$ as suitable C^∞-functions on the upper half plane, called *nearly holomorphic modular forms* of weight k and order $\leq r-1$. Using this interpretation, the Gauss-Manin connection takes the form of the so called *Maass-Shimura differential operator* $\delta_k(f) = \frac{1}{2\pi i}\left(\frac{\partial f}{\partial \tau} + \frac{k}{2iy} f\right)$ where τ is the standard coordinate on the upper half plane and $y = \mathrm{Im}(\tau)$. For $k > 2r$ one also has a *holomorphic projection* $\mathrm{H}^{\mathrm{hol}}$ to weight k modular forms and, hence, a q-expansion of nearly holomorphic forms. See Urban [2014, §2] for details. This is used, for example, to study special values of triple product L-functions as follows.

Let f, g, h be a triple of normalized primitive cuspidal classical eigenforms of weights k, ℓ, m, characters χ_f, χ_g, χ_h and tame levels N_f, N_g, N_h respectively. We write $f \in S_k(N_f, \chi_f)$, $g \in S_\ell(N_g, \chi_g)$, $h \in S_m(N_h, \chi_h)$. We assume that (k, ℓ, m) is unbalanced, i.e., there is an integer $t \geq 0$ such that $k - \ell - m = 2t$. We set $N := \ell.\mathrm{c.m.}(N_f, N_g, N_h)$ and $\mathbb{Q}_{f,g,h} := \mathbb{Q}_f \cdot \mathbb{Q}_g \cdot \mathbb{Q}_h$ the number field generated over $\mathbb{Q}$ by the Hecke eigenvalues of f, g, h. We assume that $\chi_f \cdot \chi_g \cdot \chi_h = 1$.

A result of Harris and Kudla [1991], previously conjectured by H. Jacquet and recently refined by Ichino [2008] and Watson [2002] implies that there are choices of Hecke-equivariant embeddings of $S_k(N_f, \chi_f)$, $S_\ell(N_g, \chi_g)$, $(S_m(N_h, \chi_h)$ into $S_k(N)$, $S_\ell(N)$, $S_m(N)$ respectively such that the images f^o, g^o, h^o of f, g, h respectively satisfy Ichino's formula, i.e.,

$$L\left(f, g, h, \frac{k+\ell+m-2}{2}\right) = (\text{non-zero algebraic constant}) \times |I(f^o, g^o, h^o)|^2,$$

where

$$I(f^o, g^o, h^o) := \frac{\langle (f^o)^*, \mathrm{H}^{\mathrm{hol}}\big(\delta^t (g^o)^{[p]} \times h^o\big)\rangle}{\langle (f^o)^*, (f^o)^*\rangle}.$$

Here $L(f,g,h,s)$ is the complex Garrett-Rankin triple product L-function attached to f,g,h. We have denoted by $\langle\ ,\ \rangle$ the Peterson inner product on the space of weight k-modular forms, $(f^o)^* = f^o \otimes \chi_f^{-1}$, $(g^o)^{[p]}$ is defined on q-expansions by: $(g^o)^{[p]}(q) := \sum_{n=1,(n,p)=1}^{\infty} a_n q^n$ if $g^o(q) = \sum_{n=1}^{\infty} a_n q^n$ and finally δ, $\mathrm{H}^{\mathrm{hol}}$ are the operators on nearly holomorphic forms introduced above.

5.1 p-Adic variations of $(\mathrm{H}_E, \mathrm{Fil}_\bullet)$. The task before us is to p-adically interpolate the constructions over the complex numbers previously described. We fix $I := [0,b]$ a closed interval, with $b \in \mathbb{Z}_{>0}$ and let $\mathcal{W}_I$ be the open adic subspace of $\mathcal{W}$ defined by

$$\mathcal{W}_I := \{x \in \mathcal{W} \quad | \quad |T^b|_x \leq |p|_x \neq 0\}.$$

Let $r \geq 0$ be an integer and denote by $\mathcal{X}_{r,I}$ the open adic subspace of $\mathcal{X} \times \mathcal{W}_I$ defined as in Section 3 by the valuations x such that $|\widetilde{\mathrm{Ha}}^{p^{r+1}}|_x \geq |p|_x$. If $E_{r,I}$ is the inverse image of the universal generalized elliptic curve over $\mathcal{X}$. We remark that the universal character k^{un} of $\mathcal{W}_I$ is r-analytic and there is a canonical subgroup $H_r \subset E_{r,I}[p^r]$ of order p^r over $\mathcal{X}_{r,I}$. Let H_r^D denote the Cartier dual of H_r.

We denote by $\mathcal{IG}_{r,I} := \underline{\mathrm{Isom}}_{\mathcal{X}_{r,I}}(\mathbb{Z}/p^r\mathbb{Z}, H_r^D)$ the adic space over $\mathcal{X}_{r,I}$ of trivializations of H_r^D. Then $\mathcal{IG}_{r,I} \longrightarrow \mathcal{X}_{r,I}$ is a finite, étale and Galois cover with Galois group $(\mathbb{Z}/p^r\mathbb{Z})^\times$. We introduce the ideals:

i) Hdg, the ideal of $\mathcal{O}^+_{\mathcal{X}_{r,I}}$ locally generated by any lift of the Hasse invariant Ha modulo p.

ii) $\underline{\beta}_r$, the ideal of $\mathcal{O}^+_{\mathcal{X}_{r,I}}$ locally generated by $\dfrac{p}{\mathrm{Hdg}^{\frac{p^r-1}{p-1}}}$.

iii) $\underline{\delta}$, the ideal of $\mathcal{O}^+_{\mathcal{IG}_{r,I}}$ locally generated by a precisely defined $(p-1)$-st root of Hdg. For $p \geq 5$ one considers the overconvergent modular form D of weight 1 which is a certain precisely defined $(p-1)$-st root of the Eisenstein series E_{p-1}. Then D locally generates $\underline{\delta}$.

In Section 3 we have exhibited the pair of sheaves (ω_E, ω_E^+) over $\mathcal{IG}_{r,I}$ which are invertible $\mathcal{O}_{\mathcal{IG}_{r,I}}$ and respectively $\mathcal{O}^+_{\mathcal{IG}_{r,I}}$-modules and the modification $\omega_E^\#$ of ω_E^+, an $\mathcal{O}^+_{\mathcal{IG}_{r,I}}$ submodule of ω_E^+ which is itself invertible. In fact in the $g=1$ case, the situation is very simple and we happen to have $\omega_E^\# = \underline{\delta} \cdot \omega_E^+$, which implies that over $\mathcal{IG}_{r,I}$ for $p \geq 5$ we have $\omega_E^\# = D \cdot \mathcal{O}^+_{\mathcal{IG}_{r,I}}$, i.e., it is globally free.

Moreover if $\psi : \mathbb{Z}/p^r\mathbb{Z} \cong H_r^D$ denotes the universal trivialization of H_r^D over $\mathcal{IG}_{r,I}$ then $P := \psi(1)$ is a universal generator of H_r^D over $\mathcal{IG}_{r,I}$ and $s := \mathrm{HT}(P)$ is a $\mathcal{O}^+_{\mathcal{IG}_{r,I}}/\underline{\beta}_r$-basis of $\omega_E^\#/\underline{\beta}_n\omega_E^\#$. In other words the pair $(\omega_E^\#, s)$ is a locally free sheaf with a marked section as in Section 2.

Let us now denote by H_E^+ the locally free $\mathcal{O}^+_{\mathfrak{X}_{r,I}}$-module of rank 2 characterized by the following property. For $U = \mathrm{Spa}(B, B^+) \subset \mathfrak{X}_{r,I}$ an open such that the generalized elliptic curve E/B is in fact defined over B^+ we let $\pi : E \longrightarrow \mathrm{Spf}(B^+)$ be the structural morphism. Then

$$\mathrm{H}_E^+|_U = \mathbb{R}^1\pi_*\big(\Omega^\bullet_{E/B^+}(\log(\pi^{-1}(\mathrm{cusps})))\big).$$

We have $\mathrm{H}_E^+ \otimes_{\mathcal{O}^+_{\mathfrak{X}_{r,I}}} \mathcal{O}_{\mathfrak{X}_{r,I}} = \mathrm{H}_E$ and H_E^+ has a natural Hodge filtration $\mathrm{Fil}^+_\bullet$, expressed by the exact sequence:

$$0 \longrightarrow \omega_E^+ \longrightarrow \mathrm{H_E}^+ \longrightarrow (\omega_E^+)^{-1} \longrightarrow 0.$$

We also have a connection on H_E^+ but we will discuss it later.

It is natural to consider: $\mathrm{H}_E^\# := \underline{\delta}\cdot\mathrm{H}_E^+$ and $\mathrm{Fil}^\#_\bullet := \underline{\delta}\cdot\mathrm{Fil}^+_\bullet$ as $\big(\mathrm{H}_E^\#, s = \mathrm{HT}(P)\big)$ is a pair consisting of a locally free sheaf of rank 2 with a marked section and $(\omega_E^\#, s) \subset (\mathrm{H}_E^\#, s)$ is a compatible filtration. Let us then consider the sequence of adic spaces and morphisms

$$\mathcal{T}_{\underline{\beta}_n}(\mathrm{H}_E^\#, s) \xrightarrow{u} \mathfrak{IG}_{r,I} \xrightarrow{v} \mathfrak{X}_{r,I}$$

and denote by $\rho := v \circ u$. Here $\mathcal{T}_{\underline{\beta}_r}(\mathrm{H}_E^\#, s)$ denotes the VBMS of Section 2 associated to the pair $(\mathrm{H}_E^\#, s)$ and ideal sheaf $\underline{\beta}_n$. This VBMS was denoted $\mathbb{V}_0(\mathrm{H}_E^\#, s)$ in Andreatta and Iovita [2017]. Then we have a natural action of $\mathbb{Z}_p^*$ on the sheaf $\mathbb{W}^+ := \rho_*\big(\mathcal{O}^+_{\mathbb{V}_0(\mathrm{H}_E^\#, s)}\big)$.

Definition 5.1.1. *We denote by $k : \mathbb{Z}_p^\times \longrightarrow \Lambda_I^\times$ a weight in $\mathfrak{W}_I$ (it could be the universal weight or not), denote by $\mathbb{W}_k^+$ the $\mathcal{O}^+_{\mathfrak{X}_{r,I}}$-module $\mathbb{W}^+[k]$, i.e. $\mathbb{W}_k^+$ is the sub-$\mathcal{O}^+_{\mathfrak{X}_{r,I}}$ module of sections of $\mathbb{W}^+$ on which $\mathbb{Z}_p^\times$ acts by multiplication with the values of k. The formalism of vector bundles with marked sections implies that $\mathbb{W}_k^+$ has a filtration by locally free, coherent $\mathcal{O}^+_{\mathfrak{X}_{r,I}}$-submodules $\mathrm{Fil}^+_{k,\bullet}$.*

We let $\mathbb{W}_k := \mathbb{W}_k^+ \otimes_{\mathcal{O}^+_{\mathfrak{X}_{r,I}}} \mathcal{O}_{\mathfrak{X}_{r,I}}$. It is a sheaf of Banach modules on $\mathfrak{X}_{r,I}$ with a filtration $\mathrm{Fil}_{k,\bullet}$ and $\mathrm{Fil}_{k,0}$ coincides with the sheaf $\omega^k_{r,I}$ of Section 3.2.

5.2 p-Adic variations of the connection. In order to obtain a connection on $\mathbb{W}_k$ we need to first choose a formal model of the morphism $\pi : E \longrightarrow \mathfrak{X}_{r,I}$, say $\tau : \mathcal{E} \longrightarrow \mathfrak{X}$. Our favorite such formal model is obtained by taking for $\mathfrak{X}$ the partial blow-up of the base change of the formal completion of the modular curve $X_1(N)$ over $\mathbb{Z}_p$ to the formal weight space $\mathrm{Spf}(\Lambda_I)$, with respect to the ideal $(p, \mathrm{Hdg}^{p^{r+1}})$ and taking $\mathcal{E}$ to be the inverse image of the generalized elliptic curve over $X_1(N)$. We also obtain a natural formal model $\mathfrak{IG}$ of $\mathfrak{IG}_{r,I}$ given by the normalization of $\mathfrak{X}$ in $\mathfrak{IG}_{r,I}$. Having fixed these formal models we obtain: a canonical $\mathcal{O}^+_{\mathfrak{IG}_{r,I}}$-submodule $\Omega^{1,+}_{\mathfrak{IG}_{r,I}/\mathfrak{W}_I}(\log)$ of $\Omega^1_{\mathfrak{IG}_{r,I}/\mathfrak{W}_I}(\log)$ and a natural connection

$$\nabla^+ : \mathrm{H}^+_E \longrightarrow \mathrm{H}^+_E \otimes_{\mathcal{O}^+_{\mathfrak{IG}_{r,I}}} \Omega^{1,+}_{\mathfrak{IG}_{r,I}/\mathfrak{W}_I}(\log),$$

whose generic fiber is the connection ∇ described at the beginning of this section.

The connection ∇^+, the weight $k : \mathbb{Z}_p^\times \longrightarrow \Lambda_I^\times$ and the formalism of VBMS produce a connection

$$\nabla_k : \mathbb{W}_k \longrightarrow \mathbb{W}_k \otimes_{\mathcal{O}_{\mathfrak{X}_{r,I}}} \Omega^1_{\mathfrak{X}_{r,I}/\mathfrak{W}_I}(\log(\mathrm{cusps}))$$

whose properties are described in the next theorem.

Theorem 5.2.1. *a) The connection ∇_k satisfies Griffith's transversality property with respect to the filtration i.e. $\nabla_k(\mathrm{Fil}_{k,i}) \subset \mathrm{Fil}_{k,i+1} \otimes_{\mathcal{O}_{\mathfrak{X}_{r,I}}} \Omega^1_{\mathfrak{X}_{r,I}/\mathfrak{W}_I}(\log)$, for all $i \geq 0$.*

b) If $\alpha \in \mathbb{Z}_{>0} \cap \mathfrak{W}_I(\mathbb{Q}_p, \mathbb{Z}_p)$ then the specialization at α of $\left(\mathbb{W}_k, \mathrm{Fil}_{k,\bullet}, \nabla_k\right)$, which we denote by $\left(\mathbb{W}_\alpha, \mathrm{Fil}_{\alpha,\bullet}, \nabla_\alpha\right)$, has $\left(\mathrm{Sym}^\alpha(\mathrm{H}_E), \mathrm{Fil}_{\alpha,\bullet}, \nabla_\alpha\right)$ as canonical submodule with filtration and connection. Moreover their global sections with slopes $h < \alpha - 1$ are equal (classicity).

For every Λ_I-valued weight k of $\mathfrak{W}_I$ the elements $\mathrm{H}^0(\mathfrak{X}_{r,I}, \mathbb{W}_k)$ have natural *q-expansions* (for details see Andreatta and Iovita [ibid.].)

Another very interesting occurrence is the fact that given a Λ_I-valued weight k of $\mathfrak{W}_I$ satisfying certain conditions (see below) the integral powers $(\nabla_k)^n$ of the connection ∇_k, for all $n \in \mathbb{Z}_{>0}$ (when we write $(\nabla_k)^n$ we really mean $\nabla_{k+2(n-1)} \circ \nabla_{k+2(n-2)} \circ \ldots \circ \nabla_k$) can be interpolated p-adically on $\mathrm{H}^0(\mathfrak{X}_{r,I}, \mathbb{W}_k)^{U_p=0}$ to the expense of possibly increasing r. More precisely we have (see Andreatta and Iovita [ibid.] for more details).

Theorem 5.2.2. *For every pair of weights γ, k in $\mathfrak{W}_I$ satisfying the assumptions Andreatta and Iovita [ibid.] Assumption 4.1 there is $b \geq r$ such that for every $w \in \mathrm{H}^0\left(\mathfrak{X}_{r,I}, \mathbb{W}_k\right)^{U_p=0}$ we have a unique section $\nabla^\gamma_k(w) \in \mathrm{H}^0(\mathfrak{X}_{b,I}, \mathbb{W}_{k+2\gamma})$ satisfying the property: if the q-expansion of w is $w(q) := \sum_{n=0}^\infty a_n q^n$ then the q-expansion of $\nabla^\gamma_k(w)$ is $\nabla^\gamma_k(w)(q) := \sum_{n=1,(p,n)=1}^\infty \gamma(n) a_n q^n$.*

5.3 The overconvergent projection. Finally, in view of the applications to the triple product p-adic L-functions which we have in mind, we define the "overconvergent projection" which is seen as a p-adic analogue of Shimura's "holomorphic projection".

Let us fix a Λ_I-valued weight k of $\mathfrak{W}_I$ and denote by $\mathbb{W}^\bullet_k$ the complex of sheaves $\mathbb{W}_k \xrightarrow{\nabla_k} \mathbb{W}_k \otimes_{\mathcal{O}_{\mathfrak{X}_{r,I}}} \Omega^1_{\mathfrak{X}_{r,I}/\mathfrak{W}_I}$ on $\mathfrak{X}_{r,I}$. We denote by $\mathrm{H}^i_{\mathrm{dR}}(\mathfrak{X}_{r,I}, \mathbb{W}^\bullet_k)$ for $i \geq 0$, the i-th hypercohomology group with values in the complex $\mathbb{W}^\bullet_k$.

We have natural actions of all the Hecke operators on these cohomology groups and remark that if $h \geq 0$ is a finite slope, we have natural slope decompositions for the action of the operator U_p of the groups $\mathrm{H}^i_{\mathrm{dR}}(\mathfrak{X}_{r,I}, \mathbb{W}_k^\bullet)$ and we denote by $\mathrm{H}^i_{\mathrm{dR}}(\mathfrak{X}_{r,I}, \mathbb{W}_k^\bullet)^{\leq h}$ the subgroup of slope less then or equal to h classes for the action of U_p (see Andreatta and Iovita [2017] section §3.8). If we denote by $\mathfrak{K}$ the total ring of fractions of Λ_I, we can describe the base change of $\mathrm{H}^1_{\mathrm{dR}}(\mathfrak{X}_{r,I}, \mathbb{W}_k^\bullet)^{\leq h}$ to $\mathfrak{K}$ as follows:

$$\mathrm{H}^1_{\mathrm{dR}}(\mathfrak{X}_{r,I}, \mathbb{W}_k^\bullet)^{\leq h} \otimes_{\Lambda_I[1/p]} \mathfrak{K} \cong \mathrm{H}^0(\mathfrak{X}_{r,I}, \omega_{r,I}^{k+2})^{\leq h} \otimes_{\Lambda_I[1/p]} \mathfrak{K}.$$

Therefore the "overconvergent projection" denoted $H^\dagger$ is the natural map obtained as the composition:

$$\mathrm{H}^0(\mathfrak{X}_{r,I}, \mathbb{W}_k)^{\leq h} \longrightarrow \mathrm{H}^1_{\mathrm{dR}}(\mathfrak{X}_{r,I}, \mathbb{W}_k^\bullet)^{\leq h} \otimes_{\Lambda_I[1/p]} \mathfrak{K} \cong \mathrm{H}^0(\mathfrak{X}_{r,I}, \omega_{r,I}^{k+2})^{\leq h} \otimes_{\Lambda[1/p]} \mathfrak{K}.$$

5.4 Application: the triple product p-adic L-function in the finite slope case. Let f, g, h be a triple of normalized primitive cuspidal classical eigenforms of weights k, ℓ, m, characters χ_f, χ_g, χ_h and tame levels N_f, N_g, N_h respectively. Write f^o, g^o, h^o for their images in $S_k(N), S_\ell(N), S_m(N)$ respectively as explained at the beginning of this section. We assume that f has finite slope a and that (k, ℓ, m) is unbalanced, i.e., there is an integer $t \geq 0$ such that $k - \ell - m = 2t$. We denote by K a finite extension of $\mathbb{Q}_p$ which contains all the values of χ_f, χ_g, χ_h. Let $\alpha_f, \alpha_g, \alpha_h$ denote overconvergent families of modular forms interpolating f^o, g^o, h^o in weights k, ℓ, m respectively. More precisely there are: a non-negative integer r, closed intervals I_f, I_g and I_h such that the weights of these families, denoted respectively $k_f : \mathbb{Z}_p^\times \to \Lambda_{I_f,K}^\times$, $k_g : \mathbb{Z}_p^\times \to \Lambda_{I_g,K}^\times$, $k_h : \mathbb{Z}_p^\times \to \Lambda_{I_h,K}^\times$ are all adapted to a certain integer $n \geq 0$. This data gives an adic space $\mathfrak{X}_{r,I} \longrightarrow \mathfrak{X}$, where I is a closed interval containing $I_f \times I_g \times I_h$.

We denote by $\omega^{k_f}, \omega^{k_g}, \omega^{k_h}$ the respective modular sheaves (over $\mathfrak{X}_{r,I}$), then $\alpha_f \in \mathrm{H}^0(\mathfrak{X}_{r,I_f}, \omega^{k_f})$, $\alpha_g \in \mathrm{H}^0(\mathfrak{X}_{r,I_g}, \omega^{k_g})$, $\alpha_h \in \mathrm{H}^0(\mathfrak{X}_{r,I_h}, \omega^{k_h})$. We make the following assumption on the weights of $\alpha_f, \alpha_g, \alpha_h$:

1) Suppose that the weights k_f, k_g, k_h are such that $k_f - k_g - k_h$ is even, i.e., there is a weight $u : \mathbb{Z}_p^\times \longrightarrow (\Lambda_{I,K})^\times$ with $2u = k_f - k_g - k_h$.

2) the weights k_g, u are each of the form: a finite order character multiplied a strongly analytic weight (see Andreatta and Iovita [ibid.]).

We see $\alpha_f, \alpha_g, \alpha_h$ as global sections of $\mathrm{Fil}_0(\mathbb{W}_{k_f}^{\mathrm{an}})$, $\mathrm{Fil}_0(\mathbb{W}_{k_g}^{\mathrm{an}})$ and $\mathrm{Fil}_0(\mathbb{W}_{k_h}^{\mathrm{an}})$ respectively. In particular we have that $(\nabla_{k_g})^u(\alpha_g^{[p]})$ makes sense and

$$(\nabla_{k_g})^u(\alpha_g^{[p]}) \in \mathrm{H}^0(\mathfrak{X}_{r',I}, \mathbb{W}_{k_g+2u}^{\mathrm{an}}),$$

for some positive integer $r' \geq r$. Therefore we have a section

$$(\nabla_{k_g})^u\big(\alpha_g^{[p]}\big) \times \alpha_h \in \mathrm{H}^0(\mathcal{X}_{r',I_u}, \mathbb{W}^{\mathrm{an}}_{k_f}).$$

Consider its class in $\mathrm{H}^1\big(\mathcal{X}_{r',I_u}, \omega^{k_f-2}\big) \otimes \mathfrak{K}_f$ via the natural projection and its overconvergent projection

$$H^\dagger\Big(\big(\nabla_{k_g})^u\big(\alpha_g^{[p]}\big) \times \alpha_h\Big) \in \mathrm{H}^0\big(\mathcal{X}_{r',I_u}, \omega^{k_f}\big) \otimes_{\Lambda_{I_f}} \mathfrak{K}_f,$$

to which we can apply the slope smaller or equal to a projector, $e^{\leq a}$:

$$e^{\leq a}\Big(H^\dagger\Big(\big(\nabla_{k_g})^u\big(\alpha_g^{[p]}\big) \times \alpha_h\Big)\Big) \in \mathrm{H}^0\big(\mathcal{X}_{r',I_u}, \omega^{k_f}\big)^{\leq a} \otimes_{\Lambda_{I_f}} \mathfrak{K}_f.$$

We are finally able to define the Garrett-Rankin triple product p-adic L-function attached to the triple $(\alpha_f, \alpha_g, \alpha_h)$ of p-adic families of modular forms, of which α_f has finite slope $\leq a$, to be:

$$\mathcal{L}_p^f(\alpha_f, \alpha_g, \alpha_h) := \frac{\Big\langle \alpha_f^*, e^{\leq a}\Big(H^\dagger\Big(\big(\nabla_{k_g})^u\big(\alpha_g^{[p]}\big) \times \alpha_h\Big)\Big)\Big\rangle}{\langle \alpha_f^*, \alpha_f^*\rangle} \in \mathfrak{K}_f \widehat{\otimes} \Lambda_{k_g,K} \widehat{\otimes} \Lambda_{k_h,K}.$$

By the definition of the overconvergent projection the p-adic L-function $\mathcal{L}_p^f(\alpha_f, \alpha_g, \alpha_h)$ has only finitely many poles, i.e., it is meromorphic.

Remark 5.4.1. The triple product p-adic L-function attached to a triple of **ordinary** families of modular forms has been defined in Darmon and Rotger [2014], using q-expansions.

Let now $x \in \mathcal{W}_{I_f}$, $y \in \mathcal{W}_{I_g}$, $z \in \mathcal{W}_{I_g}$ be a triple of unbalanced classical weights, i.e., such that $x, y, z \in \mathbb{Z}_{\geq 2}$ and such that there is $t \in \mathbb{Z}_{\geq 0}$ with $x - y - z = 2t$. Let us denote by f_x, g_y, h_z the specializations of $\alpha_f, \alpha_g, \alpha_h$ at x, y, z respectively, seen as sections over $\mathcal{X}_{r',I_u}$ of $\omega^x \subset \mathrm{Fil}_{x-2}(\mathbb{W}^{\mathrm{an}}_{x-2}) = \mathrm{Symm}^{x-2}(\mathrm{H}_E)$, $\omega^y \subset \mathrm{Fil}_{y-2}(\mathbb{W}^{\mathrm{an}}_{y-2}) = \mathrm{Symm}^{y-2}(\mathrm{H}_E)$, $\omega^z \subset \mathrm{Fil}_{z-2}(\mathbb{W}^{\mathrm{an}}_{z-2}) = \mathrm{Symm}^{z-2}(\mathrm{H}_E)$ respectively.

If we fix embeddings of $\overline{\mathbb{Q}}$ in $\mathbb{C}$ and $\mathbb{C}_p$ respectively, using the identifications between the p-adic overconvergent projection and the complex holomorphic one and between the Gauss-Manin connection and the Shimura-Maass differential operator on the one hand and the classical expressions of the special values of the complex triple product L-functions on the other, we obtain:

$$|\mathcal{L}_p^f(\alpha_f, \alpha_g, \alpha_h)(x, y, z)|_p = (\text{explicit constant}) \times \Big(\mathrm{L}^{\mathrm{alg}}(f_x, g_y, h_z, \frac{x+y+z-2}{2})\Big)^{\frac{1}{2}}.$$

In particular for $x = k, y = \ell, z = m$ we have $\mathcal{L}_p^f(\alpha_f, \alpha_g, \alpha_h)(k, \ell, m) \neq 0$ which implies that $\mathcal{L}_p^f(\alpha_f, \alpha_g, \alpha_h) \neq 0$.

6 Higher coherent cohomology

The purpose of this last section is to explain how the higher coherent cohomology of automorphic bundles enters the picture and how this is related to irregular motives. Let $K \subset \mathrm{GSp}_{2g}(\mathbb{A}_f)$ be a compact open subgroup. Let $X_K \to \operatorname{Spec} \mathbb{C}$ be a toroidal compactification of the Siegel variety of genus g and level K. For any classical weight $k = (k_1, \cdots, k_g)$, we can consider the cuspidal cohomology $\mathrm{H}^i(X_K, \omega^k(-D))$, as well as the usual cohomology $\mathrm{H}^i(X_K, \omega^k)$. They don't depend on the choice of the toroidal compactification. Let us define the interior coherent cohomology

$$\overline{\mathrm{H}}^i(X_K, \omega^k) = \mathrm{Im}(\mathrm{H}^i(X_K, \omega^k(-D)) \to \mathrm{H}^i(X_K, \omega^k))$$

Recall that $\lim_K \overline{\mathrm{H}}^i(X_K, \omega^k)$ is an admissible $\mathrm{GSp}_{2g}(\mathbb{A}_f)$-representation. We first recall the following result, saying that a generic weight has only cohomology in one degree.

Theorem 6.0.1 (Harris [1990], Li and Schwermer [2004], Lan [2016]). *There is an (explicit) constant C[6] which only depends on g such that if $(k = (k_1, \cdots, k_g))$ and:*

1. *$|k_i - k_{i+1}| \geq C$ for all $1 \leq i \leq g-1$,*
2. *$|k_i + k_j| \geq C$ for all $1 \leq i \leq j \leq g$*

then $\overline{\mathrm{H}}^\star(X_K, \omega^k)$ is concentrated in one degree.

Let us explain how one should think about this theorem. Identify $\mathbb{Z}^g$ with the space of characters of the maximal diagonal torus of the group Sp_{2g}. We make a choice of positive roots R^+ to be the union of the compact positive roots $R_c^+ = \{e_i - e_j\}_{1 \leq i < j \leq g}$ and non-compact positive roots $R_{nc}^+ = \{e_i + e_j\}_{1 \leq i \leq j \leq g}$.

We can associate to k the g-uple $\lambda = (\lambda_1, \cdots, \lambda_g) = (k_1 - 1, \cdots, k_g - g) = k + \rho_c - \rho_{nc} \in \mathbb{Z}^g$ where ρ_c is half the sum of the positive compact roots, and ρ_{nc} is half the sum of the positive non-compact roots. We see that k is dominant if and only if λ is R_c^+ regular: $\langle \lambda, \alpha \rangle > 0$ for all $\alpha \in R_c^+$.

The theorem above says that if k is such that λ is far enough from all the walls perpendicular to all the roots, then $\overline{\mathrm{H}}^\star(X, \omega^k)$ is concentrated in one single degree which can be determined as follows: let $C \subset \mathbb{Z}^g$ be the chamber defined by $\lambda_1 > \ldots > \lambda_g \geq 0$; the cohomological degree is the minimum of the length of the elements of the Weyl group $W_{Sp_{2g}} = (\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ that take λ to an element of C. Let λ be far enough from the

[6] One can be more precise. For example, if $g = 1$, $\overline{\mathrm{H}}^\star(X_K, \omega^k)$ is concentrated in degree 0 if $k \geq 2$ and in degree 1 if $k \leq 0$. If $g = 2$, $\overline{\mathrm{H}}^i(X_K, \omega^k)$ is concentrated in degree 0 if $k_2 \geq 4$; degree 1 if $k_2 \leq 0$ and $k_1 + k_2 \geq 5$; degree 2 if $k_2 + k_1 \leq 1$ and $k_1 \geq 3$; degree 3 if $k_1 \leq -1$.

walls and let w be an element of the Weyl group of minimal length such that $w.\lambda \in C$. Although the Hecke modules $\overline{\mathrm{H}}^{\ell(w)}(X_K, \omega^{\lambda-\rho_c+\rho_{nc}})$ and $\overline{\mathrm{H}}^0(X_{\mathbb{C}}, \omega^{w.\lambda-\rho_c+\rho_{nc}})$ are rarely isomorphic (except for $g = 1$), they are closely related [7].

So from that perspective, a generic weight has cohomology in one single degree, and moreover, one can often reduce to degree 0 cohomology. In that sense, Theorem 3.1 is optimal as long as we want to work over the total weight space.

We'd now like to consider "singular" weights λ that lie on the walls $\langle\lambda, \alpha\rangle = 0$ for $\alpha \in R^+_{nc}$. The main reason is that the corresponding cohomology groups of weight $\lambda - \rho_c + \rho_n$ are conjecturally related to irregular motives. Moreover, they don't admit a Betti cohomology realization and can only be seen in the coherent cohomology. In one direction, one knows how to attach compatible systems of Galois representations to automorphic forms realized in the coherent cohomology [8] (Deligne and Serre [1974], Taylor [1991], Goldring [2014], Pilloni and Stroh [2016], Boxer [2015], Goldring and Koskivirta [2017]). The method is to establish congruences with automorphic forms which are holomorphic discrete series at infinity and whose Galois representations can (often) be constructed in the étale cohomology of a Shimura variety.

Example 1 (Limits of discrete series). Let π be an automorphic representation for the group $\mathrm{GSp}_{2g}/\mathbb{Q}$ for which π_∞ is a limit of discrete series with infinitesimal character λ lying on such non-compact wall. Then π_f is realized in $\lim_K \overline{\mathrm{H}}^i(X_K, \omega^{\lambda-\rho_c+\rho_{nc}})$. Moreover, it will often be realized (for instance if the associated parameter has trivial centralizer) in several consecutive degrees (the number of consecutive degrees is the number of non-compact roots $\alpha \in R^+_{nc}$ such that $\langle\lambda, \alpha\rangle = 0$). For the standard $2g+1$ dimensional representation of the L-group $\mathrm{GSp}^L_{2g} \to \mathrm{GL}_{2g+1}$, the associated compatible system has (conjectural) Hodge-Tate weights $(\lambda_1, -\lambda_1, \cdots, \lambda_g, -\lambda_g, 0)$. The simplest situation is $g = 1$, $\lambda = 0$, $k = 1$. There is an isomorphism of Hecke modules $\overline{\mathrm{H}}^0(X_K, \omega^1) = \overline{\mathrm{H}}^1(X_K, \omega^1)$ and $\lim_K \overline{\mathrm{H}}^i(X_K, \omega^1) = \oplus \pi_f$ where π_f runs over all admissible $\mathrm{GL}_2(\mathbb{A}_f)$-modules for which $\pi_\infty \otimes \pi_f$ is cuspidal automorphic for π_∞ the unique limit of discrete series of $\mathrm{GL}_2(\mathbb{R})$ [9]. For the tautological 2-dimensional representation of the L-group, the associated compatible system arises from an Artin motive (Deligne and Serre [1974]).

When $g = 1$ we have the following theorem (a particular case of Artin's conjecture):

[7] If λ is far enough from the walls, all the cohomology is represented by automorphic forms $\pi_\infty \otimes \pi_f$ which are discrete series at infinity. The L-packet corresponding to π_∞ contains a holomorphic discrete series π^h_∞. It the global Langlands parameter associated to π has trivial centralizer, then $\pi^h_\infty \otimes \pi_f$ is still automorphic and realized in the degree 0 cohomology

[8] For automorphic forms which are not holomorphic limits of discrete series, they are very "weak" compatible systems since they are not known to be de Rham.

[9] π is normalized by imposing that the central character of is $|.| \times \chi$ where χ is a finite character

Theorem 6.0.2 (Langlands [1980], Tunnell [1981], Buzzard, Dickinson, Shepherd-Barron, and Taylor [2001], Khare and Wintenberger [2009], Kisin [2009], Kassaei [2013], Kassaei, Sasaki, and Tian [2014], Pilloni and Stroh [2016], Calegari and Geraghty [2018]**).** *There is a bijective correspondence between isomorphism classes of continuous irreducible odd Galois representations* $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$ *and cuspidal automorphic forms* $\pi = \pi_\infty \otimes \pi_f$ *on* $\mathrm{GL}_2/\mathbb{Q}$ *such that* π_∞ *is a limit of discrete series. This bijection satisfies* $L(\rho, s) = L(\pi \otimes |\,|^{-\frac{1}{2}}, s)$.

Remark 6.0.3. The theorem holds also for totally odd irreducible two dimensional representations of the Galois group G_F of a totally real finite field extension F of $\mathbb{Q}$. Under mild technical hypothesis one can also prove that a representation $\rho : G_F \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ which is totally odd, irreducible and geometric with Hodge-Tate weights all equal to 0 is an Artin representation. See Pilloni and Stroh [2016].

The case $g = 2$ is also particularly interesting. Let $A \to \operatorname{Spec} \mathbb{Q}$ be a simple abelian surface and denote by $\mathrm{H}^1(A)$ the associated motive. For every prime p, we can define a Weil-Deligne representation $WD_p(\mathrm{H}^1(A)) : \mathrm{WD}_{\mathbb{Q}_p} \to \mathrm{GSp}_4(\mathbb{C})$. By Gan and Takeda [2011] there is a local L-packet $\Pi_p(A)$ whose Langlands parameter is $WD_p(\mathrm{H}^1(A)) \otimes |\,|^{\frac{3}{2}}$. This local L-packet contains at most two elements and exactly one generic element π_p^g. At the infinite place there is a local L-packet $\Pi_\infty(A)$ which consists of the two limits of discrete series (respectively holomorphic and generic) $\{\pi_\infty^h, \pi_\infty^g\}$ with infinitesimal character $\lambda = (1, 0)$[10]. We let $\Pi(A) = \prod_p \Pi_p(A) \times \Pi_\infty(A)$. The following is a particular example of Langlands's conjectures:

Conjecture 2. *The global L-packet* $\Pi(A)$ *contains a cuspidal automorphic form. As a consequence the complex L-function* $L(\mathrm{H}^1(A), s)$ *extends to an entire function over* $\mathbb{C}$ *and satisfies a functional equation as predicted in Serre [1970].*

Remark 6.0.4. 1) If $\mathrm{End}(A) \neq \mathbb{Z}$ (the GL_2-type case), the conjecture is known thanks to the works Khare and Wintenberger [2009], Kisin [2009] and Yoshida [1984].

2) $\pi_f = \otimes_p \pi_p \in \prod_p \Pi_p(A)$ is realized (with multiplicity one) in $\lim_K \overline{\mathrm{H}}^0(X_K, \omega^{(2,2)})$ provided $\pi_\infty^h \otimes \pi_f$ is automorphic, and in $\lim_K \overline{\mathrm{H}}^1(X_K, \omega^{(2,2)})$ provided $\pi_\infty^g \otimes \pi_f$ is automorphic.

3) The character formula of Labesse and Langlands [1979] describes which elements of $\Pi(A)$ should be cuspidal automorphic. If $\mathrm{End}(A) = \mathbb{Z}$, all elements of $\Pi(A)$ should be cuspidal automorphic. If $\mathrm{End}(A) \neq \mathbb{Z}$ an element $\pi_\infty \otimes \pi_p$ should be automorphic if and only if the number of non-generic representations occurring in the product is even. If $\mathrm{End}(A) = \mathbb{Z}$ we can choose the particular element of $\Pi(A)$ which is generic at all finite

[10] we normalize here $\{\pi^h, \pi^g\}$ by asking that the central character is $|\,|^2$ on the connected component of the center

places and π_∞^h at infinity. This representation has a unique line which is generated by a lowest weight vector at ∞ and is invariant under the paramodular group of a certain level $N(A)$ at all finite places. Therefore there should be a well determined (up to scalar) holomorphic weight $(2,2)$ cuspform of paramodular level $N(A)$ with rational Hecke eigenvalues attached to A. This is the paramodular conjecture of Brumer and Kramer [2014].

Theorem 6.0.2 and Conjecture 2 share high similarities: the Hodge-Tate weights of the motives considered have multiplicity two and the relevant automorphic forms contribute to two coherent cohomology degrees of the same automorphic sheaf. We will now explain how these singular weights behave in p-adic families: this appears to be a crucial tool in the proof of Theorem 6.0.2 and in the approaches to Conjecture 2.

6.1 Modular curves and weight one forms. We slightly change notations. Let p be a prime integer and $N \geq 3$ an integer prime to p. Let X be the modular curve of level $\Gamma_1(N)$ over Spec $\mathbb{Z}_p$ and X_{Iw} the modular curve of level $\Gamma_1(N) \cap \Gamma_0(p)$. We now examine the behaviour of p-adic families at weight one. We restrict ourselves to ordinary families because weight one modular forms of finite slope at p are necessarily ordinary.

Theorem 6.1.1 (Hida [1986]). *There is a finite projective $\Lambda = \mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]$-module M* [11] *such that for all $k \in \mathbb{Z}_{\geq 2}$,*

$$M \otimes_{\Lambda,k} \mathbb{Z}_p = \mathrm{ordH}^0(X_{\mathrm{Iw}}, \omega^k(-D)).$$

Here $\mathrm{ord} = \lim_n U_p^{n!}$ is the ordinary projector for U_p. There is a control theorem in weight 1, but it is more complicated to state. By construction of M, there is an injective map $\mathrm{H}^0(X_{Iw}, \omega(-D)) \to M \otimes_{\Lambda,1} \mathbb{Z}_p$. In order to state the classicity theorem in weight 1, we need to look at the Galois representation picture. For any $k \in \mathbb{Z}$, and any eigenform f in $M \otimes_{\Lambda,k} \overline{\mathbb{Z}}_p$, there is an associated two dimensional Galois representation ρ_f whose restriction to inertia at p is nearly ordinary (χ_p is the p-adic cyclotomic character):

$$\rho_f|_{I_{\mathbb{Q}_p}} \simeq \begin{pmatrix} 1 & \star \\ 0 & \chi_p^{1-k} \end{pmatrix}$$

If $k \geq 2$ the representation ρ_f is automatically de Rham which is consistent with the control theorem. If $k = 1$, the representation is de Rham if and only if it is unramified at p and the classicity theorem in weight 1 states:

Theorem 6.1.2 (Buzzard and Taylor [1999], Pilloni and Stroh [2016]). *An eigenclass $f \in M \otimes_{\Lambda,1} \overline{\mathbb{Z}}_p$ is classical if and only if the associated Galois representation is unramified at p.*

[11] Moreover, M carries an action of the Hecke algebra and the control isomorphism is Hecke equivariant.

This classicity theorem is one of the key steps towards the proof of Theorem 6.0.2 via the strategy envisionned in Buzzard and Taylor [1999][12]: establish a congruence between an icoashedral artin representation and a modular elliptic curve modulo 5, and prove a modular lifting theorem. There is a difficulty to prove a modular lifting theorem with weight one forms because congruences are obstructed (since $\mathrm{H}^1(X, \omega) \neq 0$) and the usual Taylor-Wiles method doesn't apply. The strategy is to prove the modular lifting theorem with the module of ordinary p-adic modular forms of weight one instead (that is the module $M \otimes_{\Lambda,1} \mathbb{Z}_p$, for which the usual Taylor-Wiles method applies) and then argue via this classicity theorem. In order to obtain a full proof of Theorem 6.0.2, it is necessary to combine this strategy with solvable base change, and therefore one needs an extension of theorem Theorem 6.1.2 over totally real fields (Pilloni and Stroh [2016]). Observe that Theorem 6.0.2 was first proved in full generality as a consequence of Serre's modularity conjecture. Calegari and Geraghty [2018] found a way to modify the Taylor-Wiles method in order to apply it directly to weight one forms, therefore eliminating the use of Theorem 6.1.2. This method is very promising but as its application depends on certain conjectural inputs it has not yet given a complete new proof of Theorem 6.0.2.

6.2 The group GSp_4 and potentially modular abelian surfaces. We now let $X \to \operatorname{Spec} \mathbb{Z}_p$ be the Siegel threefold of hyperspecial level at p (and some fixed level away from p). We let $X_{\mathrm{Iw}} \to X_{\mathrm{Kli}} \to X$ be the Siegel threefolds of Iwahori and Klingen level at p. As we have seen, for all weights $k = (k_1, k_2)$ with $k_1 \geq k_2$, we have an automorphic vector bundle ω^k. The tempered part (at infinity) of the cohomology $\overline{\mathrm{H}}^{\star}(X_{\mathbb{C}}, \omega^{(k_1,k_2)})$ is concentrated in degree 0 if $k_2 \geq 3$ while there is cohomology in degree 0 and 1 if $k_2 = 2$. The situation resembles that of modular curves and weight one forms, except that there are now infinitely many singular weights: all those of the form $(k, 2)$ for $k \geq 2$.

Let $\Lambda_1 = \mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]$ and $\Lambda_2 = \mathbb{Z}_p[\![(\mathbb{Z}_p^\times)^2]\!]$. We first state the main theorem of classical Hida theory:

Theorem 6.2.1 (Hida [1986], Pilloni [2011]). *There exists a finite projective Λ_2-module M such that for (k_1, k_2) with $k_1 \geq k_2 \geq 4$*

$$M \otimes_{\Lambda_2,(k_1,k_2)} \mathbb{Z}_p = \mathrm{ordH}^0(X_{\mathrm{Iw}}, \omega^{(k_1,k_2)}(-D))$$

Here ord is the ordinary projector for the operator $U_{p,1}U_{p,2}$. The bound $k_2 \geq 4$ is an accident and the expected optimal bound is $k_2 \geq 3$. It is instructive to look at the Galois representation picture. For all eigenclasses $f \in M \otimes_{\Lambda_2,(k_1,k_2)} \overline{\mathbb{Z}}_p$ there is an associated nearly ordinary Galois representation $\rho_f : G_{\mathbb{Q}} \to \mathrm{GSp}_4(\overline{\mathbb{Q}}_p)$ such that:

[12] Granting the fact that it is known in the solvable image case by automorphic methods

$$\rho_f|_{I_{\mathbb{Q}_p}} \simeq \begin{pmatrix} 1 & \star & \star & \star \\ 0 & \chi_p^{-k_2+2} & \star & \star \\ 0 & 0 & \chi_p^{-k_1+1} & \star \\ 0 & 0 & 0 & \chi_p^{-k_1-k_2+3} \end{pmatrix}$$

Such a representation is automatically geometric if $k_1 \geq k_2 \geq 3$. It is tempting to believe that an eigenclass of weight $k_1 \geq k_2 = 2$ is classical if and only if

$$\rho_f|_{I_{\mathbb{Q}_p}} \simeq \begin{pmatrix} 1 & 0 & \star & \star \\ 0 & 1 & \star & \star \\ 0 & 0 & \chi_p^{-k_1+1} & 0 \\ 0 & 0 & 0 & \chi_p^{-k_1+1} \end{pmatrix}$$

(the analogue of Theorem 6.1.2). The techniques of Theorem 6.1.2 which crucially depend on the explicit relation between q-expansion and Hecke eigenvalues don't appear to generalize to this case. In particular, it seems impossible to generalize the strategy of Buzzard and Taylor [1999] to Conjecture 2.

We now state the main theorem of Higher Hida theory which deals with the singular weights:

Theorem 6.2.2 (Pilloni [2017]). *There exists a perfect complex $M^\bullet$*[13] *of Λ_1-modules of amplitude $[0,1]$ such that for all $k \in \mathbb{Z}_{\geq 2}$:*

$$M^\bullet \otimes^L_{(k,2)} \mathbb{Q}_p = \mathrm{ord}' \mathrm{R}\Gamma\big(X_{\mathrm{Kli}}, \omega^{(k,2)}(-D) \otimes \mathbb{Q}_p\big).$$

Here ord' is the ordinary projector for the operator $U_{p,1}$. The control theorem is sharp. Let us explain briefly the construction of $M^\bullet$. Let $X \to \operatorname{Spec} \mathbb{Z}_p$ be the Siegel threefold of level prime to p. We let $\mathfrak{X}$ be the p-adic completion of X and denote by $\mathfrak{X}^{\geq i}$ the p-rank stratification on $\mathfrak{X}$. As we have seen in Section 3, over the ordinary locus $\mathfrak{X}^{\geq 2}$ we have a multiplicative subgroup of rank 2 of $G[p^\infty]$ which provides the extra structure on ω_G allowing for the interpolation property. Over $\mathfrak{X}^{\geq 1}$ there is still an extra structure as we can choose a multiplicative Barsotti-Tate group of height 1, $H_\infty \hookrightarrow G[p^\infty]$ and for such a choice we have an exact sequence

$$0 \to \omega_{G[p^\infty]/H_\infty} \to \omega_{G[p^\infty]} \to \omega_{H^\infty} \to 0$$

and the Hodge-Tate map realizes an isomorphism $T_p(H_\infty^D) \otimes \mathcal{O}_{\mathfrak{X}^{\geq 1}(p^\infty)} \to \omega_{H_\infty}$. Thus, we end up with half the extra structure we had over the ordinary locus and, this allows the interpolation of the automorphic sheaves in one direction. It is quite important to work

[13] $M^\bullet$ carries an action of the Hecke algebra and the control theorem is Hecke equivariant

over this larger base since $\mathfrak{X}^{\geq 1}$ is morally of cohomological dimension 1, while $\mathfrak{X}^{\geq 2}$ is of cohomological dimension 0[14].

The projective module M is obtained by considering the (degree 0) ordinary cohomology of an interpolation sheaf over $\mathfrak{X}^{\geq 2}$ as explained in Section 3, while the complex $M^\bullet$ is obtained by considering the ordinary cohomology of an other interpolation sheaf, whose weight is parametrized by Λ_1 over $\mathfrak{X}^{\geq 1}$.

The cohomology $M^\bullet \otimes^L_{\Lambda_1,2} \mathbb{Z}_p$ is an integral modification of $\mathrm{ord}'\mathrm{R}\Gamma\big(X_{\mathrm{Kli}}, \omega^{(2,2)}(-D)\big)$. One important property of $M^\bullet$ is that it is concentrated in two degrees, while this is not known to hold for $\mathrm{R}\Gamma\big(X, \omega^{(2,2)}(-D)\big)$. In Boxer, Calegari, Gee, and Pilloni [2018] we manage to study the Galois representation supported by $M^\bullet$ and prove under some technical assumptions that it is ordinary. As a corollary, $M^\bullet \otimes^L_{\Lambda_1,2} \mathbb{Z}_p$ can be used to construct modified Taylor-Wiles systems in the sense of Calegari and Geraghty [2018]. It is an important ingredient in the proof of the following theorem:

Theorem 6.2.3 (Boxer, Calegari, Gee, and Pilloni [2018])**.** *Let $A/\mathbb{Q}$ be an abelian surface. Then there is a finite field extension F of $\mathbb{Q}$ such that $\mathrm{H}^1(A|_F)$ is automorphic. In particular $L(\mathrm{H}^1(A), s)$ has a meromorphic continuation to $\mathbb{C}$.*

The theorem holds also when $\mathbb{Q}$ is replaced by a totally real field.

References

Fabrizio Andreatta and Adrian Iovita (2017). "Triple product p-adic L-functions associated to finite slope p-adic families of modular forms, with an appendix by Eric Urban". arXiv: 1708.02785 (cit. on pp. 279, 282–284).

Fabrizio Andreatta, Adrian Iovita, and Vincent Pilloni (2015). "p-adic families of Siegel modular cuspforms". *Ann. of Math. (2)* 181.2, pp. 623–697. MR: 3275848 (cit. on pp. 268, 272).

– (2016a). "On overconvergent Hilbert modular cusp forms". *Astérisque* 382, pp. 163–193. MR: 3581177 (cit. on p. 276).

– (2016b). "The adic, cuspidal, Hilbert eigenvarieties". *Res. Math. Sci.* 3, Paper No. 34, 36. MR: 3544971 (cit. on p. 279).

– (2018). "Le Halo Spectral". To appear in the Ann. Sci. ENS (cit. on p. 277).

John Bergdall and Robert J. Pollack (2016). "Arithmetic properties of Fredholm series for p-adic modular forms". *Proc. Lond. Math. Soc. (3)* 113.4, pp. 419–444. MR: 3556487 (cit. on p. 279).

– (2017). "Slopes of modular forms and the ghost conjecture". In: (cit. on p. 279).

[14] Precisely, for all k, $\mathrm{H}^i(\mathfrak{X}^{\geq 1}, \omega^k(-D)) = 0$ if $i > 1$ and $\mathrm{H}^i(\mathfrak{X}^{\geq 2}, \omega^k(-D)) = 0$ if $i \geq 1$

Laurent Berger and Pierre Colmez (2008). "Familles de représentations de de Rham et monodromie p-adique". *Astérisque* 319. Représentations p-adiques de groupes p-adiques. I. Représentations galoisiennes et (ϕ, Γ)-modules, pp. 303–337. MR: 2493221 (cit. on p. 278).

Stéphane Bijakowski, Vincent Pilloni, and Benoît Stroh (2016). "Classicité de formes modulaires surconvergentes". *Ann. of Math. (2)* 183.3, pp. 975–1014. MR: 3488741 (cit. on p. 276).

George A. Boxer (2015). "Torsion in the Coherent Cohomology of Shimura Varieties and Galois Representations". PhD thesis. Harvard University, p. 219. arXiv: 1507.05922. MR: 3450451 (cit. on p. 287).

George A. Boxer, Frank Calegari, Toby Gee, and Vincent Pilloni (2018). "Abelian surfaces over totally real fields are potentially modular" (cit. on p. 292).

Riccardo Brasca (2013). "p-adic modular forms of non-integral weight over Shimura curves". *Compos. Math.* 149.1, pp. 32–62. MR: 3011877 (cit. on p. 276).

– (2016). "Eigenvarieties for cuspforms over PEL type Shimura varieties with dense ordinary locus". *Canad. J. Math.* 68.6, pp. 1227–1256. MR: 3563721 (cit. on p. 276).

Armand Brumer and Kenneth Kramer (2014). "Paramodular abelian varieties of odd conductor". *Trans. Amer. Math. Soc.* 366.5, pp. 2463–2516. MR: 3165645 (cit. on p. 289).

Kevin Buzzard, Mark Dickinson, Nick Shepherd-Barron, and Richard Taylor (2001). "On icosahedral Artin representations". *Duke Math. J.* 109.2, pp. 283–318. MR: 1845181 (cit. on p. 288).

Kevin Buzzard and Toby Gee (2011). "The conjectural connection between automorphic representations and Galois representations". In: Proceedings of the LMS Durham Symposium 2011 (cit. on p. 267).

Kevin Buzzard and Richard Taylor (1999). "Companion forms and weight one forms". *Ann. of Math. (2)* 149.3, pp. 905–919. MR: 1709306 (cit. on pp. 289–291).

Frank Calegari and David Geraghty (2018). "Modularity lifting beyond the Taylor–Wiles method". *Invent. Math.* 211.1, pp. 297–433. MR: 3742760 (cit. on pp. 288, 290, 292).

Przemysław Chojecki, D. Hansen, and C. Johansson (2017). "Overconvergent modular forms and perfectoid Shimura curves". *Doc. Math.* 22, pp. 191–262. MR: 3609197 (cit. on p. 274).

Laurent Clozel (1990). "Motifs et formes automorphes: applications du principe de fonctorialité". In: *Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988)*. Vol. 10. Perspect. Math. Academic Press, Boston, MA, pp. 77–159. MR: 1044819 (cit. on p. 267).

Robert Coleman and Barry Mazur (1998). "The eigencurve". In: *Galois representations in arithmetic algebraic geometry (Durham, 1996)*. Vol. 254. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 1–113. MR: 1696469 (cit. on pp. 268, 276).

Henri Darmon and Victor Rotger (2014). "Diagonal cycles and Euler systems I: A p-adic Gross-Zagier formula". *Ann. Sci. Éc. Norm. Supér. (4)* 47.4, pp. 779–832. MR: 3250064 (cit. on p. 285).

Pierre Deligne and Jean-Pierre Serre (1974). "Formes modulaires de poids 1". *Ann. Sci. École Norm. Sup. (4)* 7, 507–530 (1975). MR: 0379379 (cit. on p. 287).

Gerd Faltings and Ching-Li Chai (1990). *Degeneration of abelian varieties*. Vol. 22. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. With an appendix by David Mumford. Springer-Verlag, Berlin, pp. xii+316. MR: 1083353 (cit. on p. 272).

Laurent Fargues (2011). "La filtration canonique des points de torsion des groupes p-divisibles". *Ann. Sci. Éc. Norm. Supér. (4)* 44.6. With collaboration of Yichao Tian, pp. 905–961. MR: 2919687 (cit. on p. 273).

Jean-Marc Fontaine and Barry Mazur (1995). "Geometric Galois representations". In: *Elliptic curves, modular forms, & Fermat's last theorem (Hong Kong, 1993)*. Ser. Number Theory, I. Int. Press, Cambridge, MA, pp. 41–78. MR: 1363495 (cit. on p. 267).

Wee Teck Gan and Shuichiro Takeda (2011). "The local Langlands conjecture for $\mathrm{GSp}(4)$". *Ann. of Math. (2)* 173.3, pp. 1841–1882. MR: 2800725 (cit. on p. 288).

Wushi Goldring (2014). "Galois representations associated to holomorphic limits of discrete series". *Compos. Math.* 150.2, pp. 191–228. MR: 3177267 (cit. on p. 287).

Wushi Goldring and Jean-Stefan Koskivirta (2017). "Strata Hasse invariants, Hecke algebras and Galois representations". Preprint (cit. on p. 287).

Harish-Chandra (1968). *Automorphic forms on semisimple Lie groups*. Notes by J. G. M. Mars. Lecture Notes in Mathematics, No. 62. Springer-Verlag, Berlin-New York, pp. x+138. MR: 0232893 (cit. on p. 267).

M. Harris (1990). "Automorphic forms and the cohomology of vector bundles on Shimura varieties". In: *Automorphic forms, Shimura varieties, and L-functions. Vol. II, Proceedings of the conference held at the University of Michigan, Ann Arbor, Michigan, July 6–16, 1988*. Ed. by Laurent Clozel and James S. Milne. Vol. 10. Perspectives in Mathematics. Academic Press, Inc., Boston, MA, pp. xii+437 (cit. on pp. 268, 286).

Michael Harris and Stephen S. Kudla (1991). "The central critical value of a triple product L-function". *Ann. of Math. (2)* 133.3, pp. 605–672. MR: 1109355 (cit. on p. 280).

Valentin Hernandez (2016). "La filtration canonique des O-modules p-divisibles". arXiv: 1611.07396 (cit. on p. 276).

Haruzo Hida (1986). "Iwasawa modules attached to congruences of cusp forms". *Ann. Sci. École Norm. Sup. (4)* 19.2, pp. 231–273. MR: 868300 (cit. on pp. 268, 289, 290).

– (2002). "Control theorems of coherent sheaves on Shimura varieties of PEL type". *J. Inst. Math. Jussieu* 1.1, pp. 1–76. MR: 1954939.

Atsushi Ichino (2008). "Trilinear forms and the central values of triple product L-functions". *Duke Math. J.* 145.2, pp. 281–307. MR: 2449948 (cit. on p. 280).

Christian Johansson and James Newton (Apr. 2016). "Extended eigenvarieties for overconvergent cohomology". arXiv: 1604.07739 (cit. on p. 279).

Owen T. R. Jones (2011). "An analogue of the BGG resolution for locally analytic principal series". *J. Number Theory* 131.9, pp. 1616–1640. MR: 2802138 (cit. on p. 270).

Payman L. Kassaei (2013). "Modularity lifting in parallel weight one". *J. Amer. Math. Soc.* 26.1, pp. 199–225. MR: 2983010 (cit. on pp. 276, 288).

Payman L. Kassaei, Shu Sasaki, and Yichao Tian (2014). "Modularity lifting results in parallel weight one and applications to the Artin conjecture: the tamely ramified case". *Forum Math. Sigma* 2, e18, 58. MR: 3294624 (cit. on p. 288).

Nicholas M. Katz (1973). "p-adic properties of modular schemes and modular forms", 69–190. Lecture Notes in Mathematics, Vol. 350. MR: 0447119 (cit. on pp. 268, 274).

Chandrashekhar Khare and Jean-Pierre Wintenberger (2009). "Serre's modularity conjecture. I". *Invent. Math.* 178.3, pp. 485–504. MR: 2551763 (cit. on p. 288).

Mark Kisin (2009). "Modularity of 2-adic Barsotti-Tate representations". *Invent. Math.* 178.3, pp. 587–634. MR: 2551765 (cit. on p. 288).

J.-P. Labesse and R. P. Langlands (1979). "L-indistinguishability for $\mathrm{SL}(2)$". *Canad. J. Math.* 31.4, pp. 726–785. MR: 540902 (cit. on p. 288).

Kai-Wen Lan (2016). "Vanishing theorems for coherent automorphic cohomology". *Res. Math. Sci.* 3, Paper No. 39, 43. MR: 3571345 (cit. on p. 286).

Robert P. Langlands (1980). *Base change for* $\mathrm{GL}(2)$. Vol. 96. Annals of Mathematics Studies. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, pp. vii+237. MR: 574808 (cit. on p. 288).

Jian-Shu Li and Joachim Schwermer (2004). "On the Eisenstein cohomology of arithmetic groups". *Duke Math. J.* 123.1, pp. 141–169. MR: 2060025 (cit. on p. 286).

Ruochuan Liu, Daqing Wan, and Liang Xiao (2017). "The eigencurve over the boundary of weight space". *Duke Math. J.* 166.9, pp. 1739–1787. MR: 3662443 (cit. on p. 279).

Barry Mazur (1989). "Deforming Galois representations". In: *Galois groups over* $\mathbf{Q}$ *(Berkeley, CA, 1987)*. Vol. 16. Math. Sci. Res. Inst. Publ. Springer, New York, pp. 385–437. MR: 1012172 (cit. on p. 268).

Vincent Pilloni (2011). "Prolongement analytique sur les variétés de Siegel". *Duke Math. J.* 157.1, pp. 167–222. MR: 2783930 (cit. on p. 290).

– (2017). "Higher coherent cohomology and p-adic modular forms of singular weight" (cit. on p. 291).

Vincent Pilloni and Benoît Stroh (2016). "Surconvergence, ramification et modularité". *Astérisque* 382, pp. 195–266. MR: 3581178 (cit. on pp. 287–290).

Peter Scholze (2015). "On torsion in the cohomology of locally symmetric varieties". *Ann. of Math. (2)* 182.3, pp. 945–1066. MR: 3418533 (cit. on p. 274).

Jean-Pierre Serre (1970). "Facteurs locaux des fonctions zêta des varietés algébriques (définitions et conjectures)". In: *Séminaire Delange-Pisot-Poitou. 11e année: 1969/70.*

Théorie des nombres. Fasc. 1: Exposés 1 à 15; Fasc. 2: Exposés 16 à 24. Secrétariat Math., Paris, p. 15. MR: 3618526 (cit. on p. 288).

Glenn Stevens (2000). *Families of modular forms with positive slopes, cours à l'Institut Henri Poincaré* (cit. on p. 269).

Richard Taylor (1991). "Galois representations associated to Siegel modular forms of low weight". *Duke Math. J.* 63.2, pp. 281–332. MR: 1115109 (cit. on p. 287).

Jerrold Tunnell (1981). "Artin's conjecture for representations of octahedral type". *Bull. Amer. Math. Soc. (N.S.)* 5.2, pp. 173–175. MR: 621884 (cit. on p. 288).

Eric Urban (2014). "Nearly overconvergent modular forms". In: *Iwasawa theory 2012*. Vol. 7. Contrib. Math. Comput. Sci. Springer, Heidelberg, pp. 401–441. MR: 3586822 (cit. on p. 280).

Thomas Crawford Watson (2002). *Rankin triple products and quantum chaos*. Thesis (Ph.D.)–Princeton University. ProQuest LLC, Ann Arbor, MI, p. 81. MR: 2703041 (cit. on p. 280).

Hiroyuki Yoshida (1984). "On Siegel modular forms obtained from theta series". *J. Reine Angew. Math.* 352, pp. 184–219. MR: 758701 (cit. on p. 288).

Received 2017-12-10.

Fabrizio Andreatta
fabrizio.andreatta@unimi.it

Adrian Iovita
adrian.iovita@concordia.ca

Vincent Pilloni
vincent.pilloni@ens-lyon.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (295–308)

PERFECTOID SPACES AND THE HOMOLOGICAL CONJECTURES

Yves André

Abstract

This is a survey of recent advances in commutative algebra, especially in mixed characteristic, obtained by using the theory of perfectoid spaces. An explanation of these techniques and a short account of the author's proof of the direct summand conjecture are included. One then portrays the progresses made with these (and related) techniques on the so-called homological conjectures.

1 The direct summand conjecture

Let R be a Noetherian (commutative) ring and S a finite ring extension, and let us consider the exact sequence of finitely generated R-modules

$$0 \to R \to S \to S/R \to 0. \tag{1-1}$$

When does this sequence split? Equivalently, when is $R \to S$ *pure*, *i.e.* remains injective after any base change? This holds for instance when $R \to S$ is flat, or when R is a normal $\mathbb{Q}$-algebra, but not in general (the embedding of $\mathbb{Q}[x, y]/(xy)$ in its normal closure gives a counter-example, since it is no longer an embedding modulo $x + y$).

The direct summand conjecture, formulated by M. Hochster around 1969, claims that (1-1) splits whenever R is regular. Hochster proved it when R contains a field Hochster [1973]. R. Heitmann proved it in dimension ≤ 3 R. C. Heitmann [2002].

Recently, the author proved it in general André [2016a]:

1.0.1 Theorem. (1-1) *splits if R is regular.*

This has many (non trivially) equivalent forms. One of them is that *every ideal of a regular ring R is contracted from every finite (or integral) extension of R*. Another (more

MSC2010: primary 13D22; secondary 13H05, 14G20.
Keywords: homological conjectures, big Cohen–Macaulay algebra, perfectoid algebra, purity.

indirect) equivalent form is the following statement, which settles a question raised by L. Gruson and M. Raynaud Raynaud and Gruson [1971, p. 1.4.3][1]:

1.0.2 Theorem. *Any integral extension of a Noetherian ring descends flatness of modules.*

We will see one more equivalent form below in the framework of the so-called homological conjectures (4.2).

2 The role of perfectoid spaces

2.1 Some heuristics. After Hochster's work Hochster [1983], it is enough to prove the direct summand conjecture in the case when R is a complete unramified local regular ring of mixed characteristic $(0, p)$ and perfect residue field k. By Cohen's structure theorem, one may thus assume $R = W(k)[[x_1, \dots, x_n]]$.

In characteristic p, all proofs of the direct summand conjecture use the Frobenius endomorphism F in some way. In mixed characteristic, $R = W(k)[[x_1, \dots, x_n]]$ carries a Frobenius-like endomorphism (acting as the canonical automorphism of $W(k)$ and sending x_i to x_i^p), which however does not extend to general finite extensions S of R. To remedy this, p-adic Hodge theory suggests to "ramify deeply", by adjoining iterated p^{th} roots of $p, x_1, \dots, x_n$. Doing this, one leaves the familiar shore of Noetherian commutative algebra for perfectoid geometry, recently introduced by P. Scholze [2012].

To begin with, $W(k)$ is replaced by the non-Noetherian complete valuation ring $\mathcal{K}^o := \widehat{W(k)[p^{\frac{1}{p^\infty}}]}$. The valuation ring $\mathcal{L}^o$ of any finite extension $\mathcal{L}$ of the field $\mathcal{K}[\frac{1}{p}]$ satisfies:
(2-1)

$$F : \mathcal{L}^o/p \overset{x \mapsto x^p}{\to} \mathcal{L}^o/p \;\; \textit{is surjective, and } \mathcal{L}^o \textit{ is } p^{\frac{1}{p^\infty}}\textit{-almost finite etale over } \mathcal{K}^o,$$

this being understood in the context of almost ring theory, introduced by G. Faltings and developped by Gabber and Ramero [2003], which gives precise meaning to "up to $p^{\frac{1}{p^\infty}}$-torsion"; for instance, $p^{\frac{1}{p^\infty}}$-almost etaleness means that $p^{\frac{1}{p^\infty}}\Omega_{\mathcal{L}^o/\mathcal{K}^o} = 0$. Actually, Gabber and Ramero [ibid.] is much more general: it deals with modules over a commutative ring up to $\mathfrak{k}$-torsion, for some idempotent ideal $\mathfrak{k}$. Going beyond the case of a valuation ideal $\mathfrak{k}$ will be crucial: beside "$p^{\frac{1}{p^\infty}}$-almost" modules, we will have to consider "$(pg)^{\frac{1}{p^\infty}}$-almost" modules for some "geometric" discriminant g.

[1] *cf.* Ohi [1996] for the equivalence. Gruson and Raynaud settled the case of a finite extension and outlined that the transition to integral extensions is not a routine exercise.

2.2 Perfectoid notions. In perfectoid geometry, one works with certain Banach[2] $\mathcal{K}$-algebras $\mathfrak{A}$. One denotes by $\mathfrak{A}^o$ the $\mathcal{K}^o$-subalgebra of power-bounded elements. One says that $\mathfrak{A}$ is *uniform* if $\mathfrak{A}^o$ is bounded, and that $\mathfrak{A}$ is *perfectoid* if it is uniform and $F : \mathfrak{A}^o/p \overset{x\mapsto x^p}{\to} \mathfrak{A}^o/p$ is surjective. An example which plays a crucial role in the sequel is $\hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \dots, x_n^{\frac{1}{p^i}}]][\frac{1}{p}]$, a deeply ramified avatar of R. Morphisms of perfectoid algebras $\mathfrak{A} \to \mathfrak{B}$ are continuous algebra homomorphisms (one then says that $\mathfrak{B}$ is a perfectoid $\mathfrak{A}$-algebra).

Perfectoid algebras enjoy three fundamental stability properties Scholze [2012]:

2.2.1 ***Tensor product.*** *If $\mathfrak{B}$ and $\mathfrak{C}$ are perfectoid $\mathfrak{A}$-algebras, so is $\mathfrak{B}\hat{\otimes}_{\mathfrak{A}}\mathfrak{C}$, and $(\mathfrak{B}\hat{\otimes}_{\mathfrak{A}}\mathfrak{C})^o$ is $p^{\frac{1}{p^\infty}}$-almost isomorphic to $\mathfrak{B}^o\hat{\otimes}_{\mathfrak{A}^o}\mathfrak{C}^o$.*

2.2.2 ***Localization.*** *The ring of functions $\mathfrak{A}\{\frac{f}{g}\}$ on the subset of the perfectoid space* $\mathrm{Spa}(\mathfrak{A}, \mathfrak{A}^o)$ *where $|f| \le |g|$ holds is perfectoid, and $\mathfrak{A}\{\frac{f}{g}\}^o$ is $p^{\frac{1}{p^\infty}}$-almost isomorphic to $\mathfrak{A}^o\langle(\frac{f'}{g'})^{\frac{1}{p^\infty}}\rangle$ for some approximations f', g' of f, g which admit iterated p^{th}-roots in $\mathfrak{A}$.*

2.2.3 ***Finite etale extension.*** *Any finite etale extension $\mathfrak{B}$ of $\mathfrak{A}$ is perfectoid, and $\mathfrak{B}^o$ is a $p^{\frac{1}{p^\infty}}$-almost finite etale extension of $\mathfrak{A}^o$.*

This generalization of 2-1 to perfectoid algebras is Faltings's "almost purity theorem" Faltings [2002] as revisited by Scholze [2012] and Kedlaya and Liu [2015].

Let us explain how the second assertion of 2.2.3 follows from the first following André [2016b, p. 3.4.2]. The idea is to reduce to the case when $\mathfrak{B}$ is Galois over $\mathfrak{A}$ with Galois group G, *i.e.* $\mathfrak{B}^G = \mathfrak{A}$ and $\mathfrak{B} \otimes_{\mathfrak{A}} \mathfrak{B} \overset{\sim}{\to} \prod_G \mathfrak{B}$. This implies $\mathfrak{B}^{oG} = \mathfrak{A}^o$. On the other hand, since $\mathfrak{B}$ is a finitely generated projective $\mathfrak{A}$-module, $\mathfrak{B} \otimes_{\mathfrak{A}} \mathfrak{B} = \mathfrak{B}\hat{\otimes}_{\mathfrak{A}}\mathfrak{B}$, and one deduces from 2.2.1 (assuming $\mathfrak{B}$ perfectoid) that $\mathfrak{B}^o\hat{\otimes}_{\mathfrak{A}^o}\mathfrak{B}^o \to \prod_G \mathfrak{B}^o$ is a $p^{\frac{1}{p^\infty}}$-almost isomorphism. To get rid of the completion, one passes modulo p^m: $\mathfrak{B}^o/p^m$ is almost Galois over $\mathfrak{A}^o/p^m$, hence almost finite etale, and a variant of Grothendieck's "équivalence remarquable" Gabber and Ramero [2003, p. 5.3.27] allows to conclude that $\mathfrak{B}^o$ it itself almost finite etale over $\mathfrak{A}^o$.

2.3 Direct summand conjecture: the case when $S[\frac{1}{p}]$ is etale over $R[\frac{1}{p}]$. Let us go back to the direct summand conjecture for $R = W(k)[[x_1, \dots, x_n]]$. The special case

[2]here and in the sequel, one can work with any perfectoid field $\mathcal{K}$ (of mixed char. $(0, p)$), - *i.e.* complete, non-discretely valued, and such that F is surjective on $\mathcal{K}^o/p$. An extensive dictionary between the langage of commutative algebra and the language of non-archimedean functional analysis is presented in André [2016b, p. 2.3.1].

when $S[\frac{1}{p}]$ is etale over $R[\frac{1}{p}]$ was settled by B. Bhatt [2014] and by K. Shimomoto [2016], using 2.2.3. Here is a slightly different account, suitable to the sequel.

Let us consider the perfectoid algebra $\mathfrak{A} = \hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \dots, x_n^{\frac{1}{p^i}}]][\frac{1}{p}]$ and notice that $\mathfrak{A}^o = \hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \dots, x_n^{\frac{1}{p^i}}]]$ is a faithfully flat extension of R. By assumption $\mathfrak{B} := S \otimes_R \mathfrak{A}$ is finite etale over $\mathfrak{A}$, hence by 2.2.3, $\mathfrak{B}^o$ is $p^{\frac{1}{p^\infty}}$-almost finite etale, hence almost pure (in the sense: almost universally injective) over $\mathfrak{A}^o$. A fortiori, (1-1) almost splits after tensoring with $\mathfrak{A}^o$. In other words, if $e \in \mathrm{Ext}^1_R(S/R, R)$ denotes the class corresponding to (1-1), then $p^{\frac{1}{p^\infty}}(e \otimes 1) = 0$ in $\mathrm{Ext}^1_R(S/R, R) \otimes_R \mathfrak{A}^o \cong \mathrm{Ext}^1_{\mathfrak{A}^o}((S \otimes_R \mathfrak{A}^o)/\mathfrak{A}^o, \mathfrak{A}^o)$. One concludes that $e = 0$ by the following general elementary lemma (applied to $M = Re$ and $\mathfrak{K} = p^{\frac{1}{p^\infty}}\mathfrak{A}^o$):

2.3.1 Lemma. *André [2016a, p. 1.1.2] Let R be a local Noetherian ring, M a finitely generated R-module, A a faithfully flat R-algebra. Let $\mathfrak{K}$ be an idempotent ideal of A such that $\mathfrak{K}.M_A = 0$ and $R \cap \mathfrak{K} \neq 0$. Then $M = 0$.*

2.4 The perfectoid Abhyankar lemma. In the general case, $S \otimes_R \mathfrak{A}$ is no longer etale over $\mathfrak{A}$: one must take into account a discriminant $g \in R$ of $S[\frac{1}{p}]$ over $R[\frac{1}{p}]$. This suggests to try to generalize 2.2.3 to *ramified* extensions of perfectoid algebras. It turns out that this is possible, provided one extracts suitable roots of g in the spirit of Abhyankar's lemma.

This leads to replace everywhere "$p^{\frac{1}{p^\infty}}$-almost" by "$(pg)^{\frac{1}{p^\infty}}$-almost", thereby extending the basic setting of almost ring theory beyond the usual situation of a non-discrete valuation ring. This also leads to introduce the notion of *almost perfectoid algebra*, where $F : \mathfrak{A}^o/p \overset{x\mapsto x^p}{\to} \mathfrak{A}^o/p$ is only assumed to be $(pg)^{\frac{1}{p^\infty}}$-almost surjective André [2016b, p. 3.5.4].

2.4.1 Theorem. *André [ibid.] Let $\mathfrak{A}$ be a perfectoid $\mathfrak{K}$-algebra, which contains a compatible system of p-power roots $g^{\frac{1}{p^j}}$ of some non-zero-divisor $g \in \mathfrak{A}^o$. Let $\mathfrak{B}'$ be a finite etale $\mathfrak{A}[\frac{1}{g}]$-algebra. Let $\mathfrak{B}$ be the integral closure of $g^{-\frac{1}{p^\infty}}\mathfrak{A}$ in $\mathfrak{B}'$, so that $\mathfrak{B}[\frac{1}{g}] = \mathfrak{B}'$.*

Then $\mathfrak{B}$ is almost perfectoid, and for any n, $\mathfrak{B}^o/p^m$ is $(pg)^{\frac{1}{p^\infty}}$-almost finite etale (hence $(pg)^{\frac{1}{p^\infty}}$-almost flat) over $\mathfrak{A}^o/p^m$.

(If $g = 1$, one may use Gabber and Ramero [2003, p. 5.3.27] again to conclude that $\mathfrak{B}^o$ is itself almost finite etale over $\mathfrak{A}^o$ and recover 2.2.3).

The basic idea is to look at the pro-system of algebras of functions $\mathfrak{A}^j := \mathfrak{A}\{\frac{p^j}{g}\}$ on complements of tubular neighborhoods of the hypersurface $g = 0$ in the perfectoid space $\mathrm{Spa}(\mathfrak{A}, \mathfrak{A}^o)$, *resp.* at the pro-system $\mathfrak{B}^j := \mathfrak{B}' \otimes_{\mathfrak{A}[\frac{1}{g}]} \mathfrak{A}\{\frac{p^j}{g}\}$. Each $\mathfrak{A}^j$ is perfectoid

(2.2.2), and each $\mathfrak{B}^j$ is finite etale over $\mathfrak{A}^j$, hence perfectoid; moreover $\mathfrak{B}^{jo}$ is $p^{\frac{1}{p^\infty}}$-almost finite etale over $\mathfrak{A}^{jo}$ by almost purity (2.2.3). One can show that $\mathfrak{B}^o$ is isomorphic to $\lim \mathfrak{B}^{jo}$, and that the latter has the asserted properties,

However, in the sequel, the identification of $(\lim \mathfrak{B}^{jo})[\frac{1}{p}]$ with the integral closure of $g^{-\frac{1}{p^\infty}}\mathfrak{A}$ in $\mathfrak{B}'$ plays no role; changing notation, we will set $\mathfrak{B} := (\lim \mathfrak{B}^{jo})[\frac{1}{p}]$, which is a uniform Banach algebra, and sktech the *proof that $\mathfrak{B}$ is almost perfectoid and that for every m, $\mathfrak{B}^o/p^m$ is $(pg)^{\frac{1}{p^\infty}}$-almost finite etale over $\mathfrak{A}^o/p^m$.*

The proof involves six steps.

1) For any $r \in \mathbb{N}[\frac{1}{p}]$, *$\lim(\mathfrak{A}^{jo}/p^r)$ is $(pg)^{\frac{1}{p^\infty}}$-almost isomorphic to $\mathfrak{A}^o/p^r$.* This is essentially Scholze's perfectoid version of Riemann's extension theorem Scholze [2015, p. II.3.1] (hint: if R denotes $\mathfrak{A}^o/p^r$ for short, $\mathfrak{A}^{jo}/p^r$ is $p^{\frac{1}{p^\infty}}$-almost isomorphic to $R^j := R[(\frac{p^j}{g})^{\frac{1}{p^\infty}}]$; the key idea is that for $j' \geq j + rp^k$, $R^{j'} \to R^j$ factors through $R^{jk} := \sum_{s\leq \frac{1}{p^k}} R(\frac{p^j}{g})^s$, so that $\lim R^j \cong \lim R^{jk}$; on the other hand, the kernel and cokernel of $R \to R^{jk}$ are killed by g raised to a power which tends to 0 when $j,k \to \infty$). Passing to the limit $r \to \infty$, it follows that $\lim \mathfrak{A}^{jo} = g^{-\frac{1}{p^\infty}}\mathfrak{A}$, and this also holds under the weaker assumption that $\mathfrak{A}$ is almost perfectoid, *cf.* André [2016b, p. 4.2.2].

2) *$\lim^1(\mathfrak{B}^{jo}/p^r)$ is $(pg)^{\frac{1}{p^\infty}}$-almost zero.* The technique is similar to the one in 1), *cf.* André [ibid., p. 4.4.1].

3) *$\lim(\mathfrak{B}^{jo}/p) \xrightarrow{F} \lim(\mathfrak{B}^{jo}/p)$ is $(pg)^{\frac{1}{p^\infty}}$-almost surjective.* Indeed, taking the limit of the exact sequence $0 \to \mathfrak{B}^{jo}/p^{\frac{p-1}{p}} \to \mathfrak{B}^{jo}/p \to \mathfrak{B}^{jo}/p^{\frac{1}{p}} \to 0$, one deduces from 2) (for $r = \frac{p-1}{p}$), that $\lim(\mathfrak{B}^{jo}/p) \to \lim(\mathfrak{B}^{jo}/p^{\frac{1}{p}})$ is $(pg)^{\frac{1}{p^\infty}}$-almost surjective; on the other hand, $\mathfrak{B}^{jo}/p^{\frac{1}{p}} \to \mathfrak{B}^{jo}/p$ is an isomorphism because $\mathfrak{B}^j$ is perfectoid.

4) *$\mathfrak{B}$ is almost perfectoid.* From 3), it suffices to show that the natural map $\mathfrak{B}^o/p = (\lim \mathfrak{B}^{jo})/p \to \lim(\mathfrak{B}^{jo}/p)$ is a $(pg)^{\frac{1}{p^\infty}}$-almost isomorphism. It is easy to see that it is injective André [ibid., p. 2.8.1], and on the other hand, the composition $\lim_{F,j}(\mathfrak{B}^{jo}/p) = \mathfrak{B}^{\flat o} \to (\lim \mathfrak{B}^{jo})/p \to \lim(\mathfrak{B}^{jo}/p)$ is almost surjective by 3).

5) *$\mathfrak{B} \to \mathfrak{B}^j$ factors through a $g^{\frac{1}{p^\infty}}$-almost isomorphism $\mathfrak{B}\{\frac{p^j}{g}\} \overset{a}{\cong} \mathfrak{B}^j$.* The factorization comes from the fact that $\mathfrak{B}^j \cong \mathfrak{B}^j\{\frac{p^j}{g}\}$, and one constructs an almost inverse to $\mathfrak{B}\{\frac{p^j}{g}\} \overset{a}{\to} \mathfrak{B}^j$ as follows (*cf.* André [ibid., p. 4.4.4]): the integral closure of $\mathfrak{A}^o$ in $\mathfrak{B}'$ maps to the integral closure $\mathfrak{A}^{jo}$ in $\mathfrak{B}^j$, which is $p^{\frac{1}{p^\infty}}$-almost $\mathfrak{B}^{jo}$ (by almost purity). Passing to the limit and inverting pg, one gets a morphism $\mathfrak{B}' \xrightarrow{\delta} \mathfrak{B}[\frac{1}{g}]$, and the sought for inverse is induced by $\delta \otimes 1_{\mathfrak{A}^j}$.

6) *for every m, $\mathfrak{B}^o/p^m$ is $(pg)^{\frac{1}{p^\infty}}$-almost finite etale over $\mathfrak{A}^o/p^m$.* This is the decisive step: how to keep track of the almost finite etaleness of $\mathfrak{B}^{jo}$ over $\mathfrak{A}^{jo}$ at the limit? As in 2.2.3, the idea is to reduce to the case when $\mathfrak{B}'$ is Galois over $\mathfrak{A}[\frac{1}{g}]$ with Galois group G, *i.e.* $(\mathfrak{B}')^G =$

$\mathfrak{A}[\frac{1}{g}]$ and $\mathfrak{B}' \otimes_{\mathfrak{A}[\frac{1}{g}]} \mathfrak{B}' \xrightarrow{\sim} \prod_G \mathfrak{B}'$. It follows that $\mathfrak{B}^j$ is G-Galois over $\mathfrak{A}^j$, and (as in 2.2.3) that $\mathfrak{B}^{jo} \hat{\otimes}_{\mathfrak{A}^{jo}} \mathfrak{B}^{jo} \to \prod_G \mathfrak{B}^{jo}$ is a $p^{\frac{1}{p^\infty}}$-almost isomorphism. Passing to the limit, (and taking into account that by 4) and 5), $\mathfrak{B} \hat{\otimes}_{\mathfrak{A}} \mathfrak{B}$ is an almost perfectoid algebra, and $\mathfrak{B} \hat{\otimes}_{\mathfrak{A}} \mathfrak{B}\{\frac{p^j}{g}\} \cong \mathfrak{B}^j \hat{\otimes}_{\mathfrak{A}^j} \mathfrak{B}^j$, so that one can apply 1)), one concludes that $\mathfrak{A}^o \to \mathfrak{B}^{oG}$ and $\mathfrak{B}^o \hat{\otimes}_{\mathfrak{A}^o} \mathfrak{B}^o \to \prod_G \mathfrak{B}^o$ are $(pg)^{\frac{1}{p^\infty}}$-almost isomorphisms. To get rid of the completion, one passes modulo p^m: $\mathfrak{B}^o/p^m$ is almost Galois over $\mathfrak{A}^o/p^m$, hence $(pg)^{\frac{1}{p^\infty}}$-almost finite etale (in constrast to 2.2.3, one cannot conclude that $\mathfrak{B}^o$ is $(pg)^{\frac{1}{p^\infty}}$-almost finite etale over $\mathfrak{A}^o$ since p may not belong to $(pg)^{\frac{1}{p^\infty}}$). □

2.5 Infinite Kummer extensions. In order to extend the strategy of 2.3 to the general case by means of the perfectoid Abhyankar lemma, one has first to adjoin to the perfectoid algebra $\hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \ldots, x_n^{\frac{1}{p^i}}]][\frac{1}{p}]$ the iterated p^{th}-roots of a discriminant g. At finite level i, it seems very difficult to control the extension

$$(W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \ldots, x_n^{\frac{1}{p^i}}]][g^{\frac{1}{p^i}}, \frac{1}{p}])^o$$

which is bigger than $W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \ldots, x_n^{\frac{1}{p^i}}]][g^{\frac{1}{p^i}}]$, but things turn easier at the infinite level, thanks to the perfectoid theory.

2.5.1 Theorem. *André [2016a, p. 2.5.2] Let $\mathfrak{A}$ be a perfectoid $\mathcal{K}$-algebra, and let $g \in \mathfrak{A}^o$ be a non-zero divisor. Then for any n, $\mathfrak{A}\langle g^{\frac{1}{p^\infty}}\rangle^o/p^m$ is $p^{\frac{1}{p^\infty}}$-almost faithfully flat over $\mathfrak{A}^o/p^m$.*

The basic idea is to add one variable T, consider the perfectoid algebra $\mathfrak{C} := \mathfrak{A}\langle T^{\frac{1}{p^\infty}}\rangle$ and look at the ind-system of algebras of functions $\mathfrak{C}_i := \mathfrak{C}\{\frac{T-g}{p^i}\}$ on tubular neighborhoods of the hypersurface $T = g$ in the perfectoid space Spa$(\mathfrak{C}, \mathfrak{C}^o)$.

The proof involves three steps.

7) *$\mathfrak{A}\langle g^{\frac{1}{p^\infty}}\rangle^o$ is $p^{\frac{1}{p^\infty}}$-almost isomorphic to $\widehat{\operatorname{colim}}_i\, \mathfrak{C}_i^o$.* This is an easy consequence of the general fact that for any uniform Banach algebra $\mathfrak{B}$ and $f \in \mathfrak{B}^o$, $\widehat{\operatorname{colim}}_i \mathfrak{B}\{\frac{f}{p^i}\}^o$ is $p^{\frac{1}{p^\infty}}$-almost isomorphic to $(\mathfrak{B}/f\mathfrak{B})^o$, *cf.* André [2016b, p. 2.9.3].

8) *$\mathfrak{C}^o$ contains a compatible system of p-power roots of some non-zero-divisor f_i such that $\mathfrak{C}_i^o$ is $p^{\frac{1}{p^\infty}}$-almost isomorphic to $\widehat{\operatorname{colim}}_j\, \mathfrak{C}\langle U\rangle^o/(p^{\frac{1}{p^j}}U - f_i^{\frac{1}{p^j}})$.* This is one instance of Scholze's approximation lemma Scholze [2012, p. 6.7]; one may assume $f_i \equiv T - g \mod p^{\frac{1}{p}}$.

9) *$\mathfrak{C}\langle U\rangle^o/(p^{\frac{1}{p^j}}U - f_i^{\frac{1}{p^j}}, p^m)$ is faithfully flat over $\mathfrak{A}^o/p^m$.* One may replace p^m by any positive power of p, *e.g.* $p^{\frac{1}{p^{j+1}}}$. Since $\mathfrak{C}$ is perfectoid, there is $g_{ij} \in \mathfrak{C}^{\frac{1}{p^\infty}}\rangle^o$ such that $g_{ij}^{p^j} \equiv$

$g \mod p$. Then $f_i^{\frac{1}{p^j}} \equiv T^{\frac{1}{p^j}} - g_{ij} \mod p^{\frac{1}{p^{j+1}}}$, and $\mathcal{C}\langle U\rangle^o/(p^{\frac{1}{p^j}}U - f_i^{\frac{1}{p^j}}, p^{\frac{1}{p^{j+1}}}) \cong (\mathfrak{A}^o/p^{\frac{1}{p^{j+1}}})[T^{\frac{1}{p^\infty}}, U]/(T^{\frac{1}{p^j}} - g_{ij})$, a free $\mathfrak{A}^o/p^{\frac{1}{p^{j+1}}}$-module. □

2.6 Conclusion of the proof of the direct summand conjecture. One chooses $g \in R$ such that $S[\frac{1}{pg}]$ is etale over $R[\frac{1}{pg}]$. One then follows the argument of 2.3, replacing $\mathfrak{A} = \hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \dots, x_n^{\frac{1}{p^i}}]][\frac{1}{p}]$ by the "infinite Kummer extension" $\mathfrak{A} := (\hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \dots, x_n^{\frac{1}{p^i}}]][\frac{1}{p}])\langle g^{\frac{1}{p^\infty}}\rangle$. One introduces the finite etale extension $\mathfrak{B}' := S \otimes_R \mathfrak{A}[\frac{1}{g}]$ of $\mathfrak{A}[\frac{1}{g}]$, and one replaces $\mathfrak{B} = S \otimes_R \mathfrak{A}$ in 2.3 by $\mathfrak{B} := (\lim \mathfrak{B}^{jo})[\frac{1}{p}]$, noting that $\mathfrak{B}^o$ is the S-algebra $\lim \mathfrak{B}^{jo}$. Translating Th. 2.4.1 and Th. 2.5.1 in terms of almost purity modulo p^m, and reasoning as in 2.3 (using the same lemma), one obtains that (1-1) splits modulo p^m for any n André [2016a, §3]. A Mittag-Leffler argument (retractions of $R/p^m \to S/p^m$ form a torsor under an artinian R/p^m-module Hochster [1973, p. 30]) shows that (1-1) itself splits. □

2.7 Derived version. In , B. Bhatt revisits this proof and proposes a variant, which differs in the analysis of the pro-system $\mathfrak{A}^{jo}/p^r$ occurring in the proof of Th. 2.4.1: he strengthens step 1) by showing that the pro-system of kernels and cokernels of $(\mathfrak{A}/p^r)_j \to (\mathfrak{A}^{jo}/p^r)_j$ is pro-isomorphic to a pro-system of $(pg)^{\frac{1}{p^\infty}}$-torsion modules; this allows to apply various functors before passing to the limit $j \to \infty$, whence a gain in flexibility. More importantly, he obtains the following derived version of the direct summand conjecture, which had been conjectured by J. de Jong:

2.7.1 Theorem. *Bhatt [n.d.] Let R be a regular Noetherian ring, and $f : X \to \operatorname{Spec} R$ be a proper surjective morphism. Then the map $R \to R\Gamma(X, \mathcal{O}_X)$ splits in the derived category $D(R)$.*

3 Existence of (big) Cohen–Macaulay algebras

3.1 Cohen–Macaulay rings and Cohen–Macaulay algebras for the non-Cohen-Macaulay rings. Let S be a local Noetherian ring with maximal ideal $\mathfrak{m}$ and residue field k. Recall that a sequence $\underline{x} = (x_1, \dots, x_n)$ in $\mathfrak{m}$ is *secant*[3] if $\dim S/(\underline{x}) = \dim S - n$, and *regular* if for every i, multiplication by x_i is injective in $S/(x_1, \dots, x_{i-1})S$. Any regular sequence

[3] following Bourkaki's terminology (for instance); it is also often called "part of a system of parameters", although $\operatorname{gr}_{\underline{x}} S$ may not be a polynomial ring in the "parameters" x_i (it *is* if $\underline{x}$ is regular).

is secant, and if the converse holds, S is said to be *Cohen–Macaulay*. Regular local rings have a secant sequence generating $\mathfrak{m}$, and are Cohen–Macaulay.

Cohen–Macaulay rings form the right setting for Serre duality and the use of local homological methods in algebraic geometry, and have many applications to algebraic combinatorics Bruns and Herzog [1993]. When confronted with a non-Cohen–Macaulay ring S, one may try two expedients:

1) *Macaulayfication*: construct a proper birational morphism $X \to \operatorname{Spec} S$ such that all local rings of X are Cohen–Macaulay. This weak resolution of singularities, introduced by Faltings, has been established in general by T. Kawasaki [2000]. Hovever, secant sequences in S may not remain secant (hence not become regular) in the local rings of X; this motivates the second approach:

2) *Construction of a Cohen–Macaulay algebra*[4]: an S-algebra C such that any secant sequence of S becomes regular in C, and $\mathfrak{m}C \neq C$.

The existence of Cohen–Macaulay algebras implies the direct summand conjecture: indeed, if C is a Cohen–Macaulay algebra for a finite extension S of a regular local ring R, it is also a Cohen–Macaulay R-algebra; this implies that $R \to C$ is faithfully flat, hence pure, and so is $R \to S$.

3.2 Constructions of Cohen–Macaulay algebras. The existence of a (big) Cohen–Macaulay algebra was established by Hochster and C. Huneke under the assumption that S contains a field Hochster and Huneke [1995]. One may assume that S is a complete local domain. In char. $p > 0$, one may then take C to be the *total integral closure of* S (*i.e.* the integral closure of S in an algebraic closure of its field of fractions). This is no longer true in the case of equal char. 0, which can nevertheless be treated by reduction to char. $p >> 0$ using ultraproduct techniques.

The remaining case of mixed characteristic was settled in André [2016a], using the same perfectoid methods, so that one has:

3.2.1 Theorem. *Any local Noetherian ring S admits a (big) Cohen–Macaulay algebra C.*

In the case of a complete local domain S of char $(0, p)$ and perfect residue field k (to which one reduces), one proceeds as follows. Cohen's theorem allows to present S as a finite extension of $R = W[[x_1, \dots, x_n]]$. One first considers the R-algebra

$$\mathfrak{A} := (\hat{\cup}_i W(k)[p^{\frac{1}{p^i}}][[x_1^{\frac{1}{p^i}}, \dots, x_n^{\frac{1}{p^i}}]][\frac{1}{p}])\langle g^{\frac{1}{p^\infty}}\rangle^o$$

[4]since it would be too restrictive to impose that C is Noetherian, one often speaks of "big" Cohen–Macaulay algebra.

and the S-algebra $\mathfrak{B}^o = \lim \mathfrak{B}^{jo}$ as above 2.6. It follows from Th. 2.4.1 and Th. 2.5.1 that $\mathfrak{B}^o$ is $(pg)^{\frac{1}{p^\infty}}$-almost isomorphic to a faithfully flat R algebra modulo any power of p. From there, one deduces that the sequence $(p, x_1, \ldots, x_n)$ is "$(pg)^{\frac{1}{p^\infty}}$-almost regular" in $\mathfrak{B}^o$.

To get rid of "almost", Lemma 2.3.1 is no longer sufficient: instead, one uses Hochster's technique of monoidal modifications Hochster [2002]Hochster and Huneke [1995]. After $\mathfrak{m}$-completion, one gets a S-algebra C in which $(p, x_1, \ldots, x_n)$, as well as any other secant sequence of S, becomes regular. □

Subsequently, using the tilting equivalence between perfectoid algebras in char. 0 and in char. p and applying Hochster's modifications in char. p rather than in char.0, K. Shimomoto [2017] shows that *in Th. 3.2.1, in mixed characteristic, C can be taken to be perfectoid. In particular, if S is regular, it admits a perfectoid faithfully flat algebra* (one may speculate about the converse).

3.3 Finite and fpqc covers. Since Cohen–Macaulay algebras for regular local rings are faithfully flat, Th. 3.2.1 implies André [2016a]:

3.3.1 Theorem. *Any finite cover of a regular scheme is dominated by a faithfully flat quasi-compact cover.*

If regularity is omitted, Spec$(\mathbb{Q}[x, y]/(xy))$ and its normalization provide a counter-example.

4 Homological conjectures

4.1 Origins from intersection theory. Under the influence of M. Auslander, D. Buchsbaum and J.-P. Serre, commutative algebra has shifted in the late 50s from the study of ideals of commutative rings to the homological study of modules (*cf.* their characterization of regular local rings by the existence of finite free resolutions for any finitely generated module, *resp.* for the residue field).

Serre proved that for any three prime ideals $\mathfrak{p}, \mathfrak{q}, \mathfrak{r}$ of a regular local ring R such that $\mathfrak{r}$ is a minimal prime of $\mathfrak{p}+\mathfrak{q}$, ht $\mathfrak{r} \leq$ ht $\mathfrak{p}+$ht $\mathfrak{q}$ Serre [1965]. The special case $\mathfrak{r} = \mathfrak{m}$ can be amplified: for any ideals I, J of R such that $I + J$ is $\mathfrak{m}$-primary, $\dim R/I + \dim R/J \leq \dim R$.

This is no longer true if R is not regular, and attempts to understand the general situation led to the so-called homological conjectures *cf.* Bruns and Herzog [1993, ch. 9], Hochster [2007].

4.2 Intersection conjectures. Let $(R, \mathfrak{m})$ be local Noetherian ring.

The first "intersection conjecture" was proposed by Peskine and Szpiro [1973], proved by them when R contains a field by reduction to char. p and Frobenius techniques, and later proved in general by P. Roberts using K-theoretic methods Roberts [1987]. It states that *if M, N are finitely generated R-modules such that $M \otimes N$ has finite length, then* $\dim N \leq \operatorname{pd} M$. It implies that *$R$ is Cohen–Macaulay if and only if there is an R-module M of finite length and finite projective dimension* (in the spirit of Serre's characterization of regular rings, which is the case $M = k$), resp. *if there is an R-module M of finite injective dimension*.

Indeed, one may take $N = R$ and deduce that $\dim R \leq \operatorname{pd} M$; by the Auslander-Buchsbaum formula, $\operatorname{pd} M = \operatorname{depth} R - \operatorname{depth} M$, so that the inequality $\operatorname{depth} R \leq \dim R$ is an equality. The second assertion follows from the fact that $\operatorname{id} M = \operatorname{depth} R$ Bass [1963].

The "new intersection conjecture", also proved by Peskine and Szpiro [1973] and Roberts [1987], states that *for any non exact complex $F_\bullet$ of free R-modules concentrated in degrees $[0, s]$ with finite length homology, $s > \dim R$.*

The "improved new intersection conjecture" is a variant due to E. Evans and P. Evans and Griffith [1981], in which the condition on $F_\bullet$ is "slightly" relaxed: the $H_{i>0}$ are of finite length and there exists a primitive cyclic submodule of H_0 of finite length. They proved it, assuming the existence of (big) Cohen–Macaulay algebras[5], and showed that it implies their "syzygy conjecture". In spite of appearances, the passage from the new intersection conjecture to its "improved" variant is no small step[6]: in fact, according to Hochster [2007] and S. Dutta [1987], the latter is *equivalent to the direct summand conjecture*.

On the other hand, in the wake of the new intersection conjecture (and motivated by the McKay correspondence in dimension 3 and the "fact" that threefold flops induce equivalences of derived categories), T. Bridgeland and S. Iyengar obtained a refinement of Serre's criterion for regular rings assuming the existence of Cohen–Macaulay algebras Bridgeland and Iyengar [2006, p. 2.4].

By Th. 1.0.1 and Th. 3.2.1, the improved new intersection conjecture and the Bridgeland-Iyengar criterion thus hold inconditionally:

4.2.1 Theorem. *Let R be a Noetherian local ring and $F_\bullet$ be a complex of finitely generated free R-modules concentrated in degree $[0, s]$, such that $H_{>0}(F_\bullet)$ has finite length.*

1. *If $H_0(F_\bullet)$ contains a cyclic R-submodule of finite length not contained in $\mathfrak{m} H_0(F_\bullet)$, then $s \geq \dim R$.*

[5] if R is Cohen–Macaulay, the Buchsbaum-Eisenbud criterion gives a condition for the exactness of $F_\bullet$ in terms of codimension of Fitting ideals of syzygies. In general, the same condition guarantees that $F_\bullet \otimes C$ is exact for any Cohen–Macaulay R-algebra C Bruns and Herzog [1993, p. 9.1.8].

[6] K-theoretic techniques failed to make the leap.

2. *If $H_0(F_\bullet)$ has finite length and contains k as a direct summand, and $s = \dim R$, then R is regular.*

And so does the syzygy conjecture:

4.2.2 Theorem. *Let R be a Noetherian local ring and M a finitely generated R-module of finite projective dimension s. Then for $i \in \{1, \dots, s-1\}$, the i-th syzygy module of M has rank $\geq i$.*

4.3 Further work around the homological conjectures using perfectoid spaces.

4.3.1 . In 2017, Heitmann and L. Ma show that Cohen–Macaulay algebras can be constructed in a way compatible with quotients $S \to S/\mathfrak{p}$ by primes of height one[7]. Using arguments similar to Bhatt's derived techniques, they deduce the vanishing conjecture for maps of Tor Hochster and Huneke [1995]:

4.3.1 Theorem. R. Heitmann and Ma [2017a] *Let $R \to S \to T$ be morphisms such that the composed map is a local morphism of mixed characteristic regular local rings, and S is a finite torsion-free extension of R. Then for every R-module M and every i, the map $\mathrm{Tor}_i^R(M, S) \to \mathrm{Tor}_i^R(M, T)$ vanishes.*

They obtain the following corollary, which generalizes results by Hochster and J. Roberts, J.-F. Boutot et al.:

4.3.2 Corollary. R. Heitmann and Ma [ibid.] *Let $R \hookrightarrow S$ be a pure, local morphism, with S regular. Then R is pseudo-rational, hence Cohen–Macaulay.*

4.3.2 . In 2017, Ma and K. Schwede define and study perfectoid multiplier/test ideals in mixed characteristic, and use them to bound symbolic powers of ideals in regular domains in terms of ordinary powers:

4.3.3 Theorem. Ma and Schwede [2017] *Let R be a regular excellent Noetherian domain and let $I \subset R$ be a radical ideal such that each minimal prime of I has height $\leq h$. Then for every n, $I^{(hn)} \subset I^n$.*

Here $I^{(hn)}$ denotes the ideal of elements of R which vanish generically to order hn at I. When R contains a field, the result was proved in Ein, Lazarsfeld, and Smith [2001] and Hochster and Huneke [1990].

[7] weak functoriality of Cohen–Macaulay algebras in general has been since announced by the author Andre [2018].

4.3.3 . An efficient and unified way of dealing with questions related to the homological conjectures in char. p is provided by "tight closure theory", which has some flavor of almost ring theory. Using Th. 2.4.1 and Th. 2.5.1 above, Heitmann and Ma give evidence that the "extended plus closure" introduced in R. C. Heitmann [2001] is a good analog of tight closure theory in mixed characteristic R. Heitmann and Ma [2017b].

References

Yves Andre (Jan. 2018). "Weak functoriality of Cohen-Macaulay algebras". arXiv: 1801.10010 (cit. on p. 305).

Yves André (2016a). "La conjecture du facteur direct". *Publications mathématiques de l'IHÉS*, pp. 1–23 (cit. on pp. 295, 298, 300–303).

– (2016b). "Le lemme d'Abhyankar perfectoïde". *Publications mathématiques de l'IHÉS*, pp. 1–70 (cit. on pp. 297–300).

Hyman Bass (1963). "On the ubiquity of Gorenstein rings". *Math. Z.* 82, pp. 8–28. MR: 0153708 (cit. on p. 304).

Bhargav Bhatt (n.d.). *On the direct summand conjecture and its derived variant*. To appear in Invent. Math (cit. on p. 301).

– (2014). "Almost direct summands". *Nagoya Math. J.* 214, pp. 195–204. MR: 3211823 (cit. on p. 298).

Tom Bridgeland and Srikanth Iyengar (2006). "A criterion for regularity of local rings". *C. R. Math. Acad. Sci. Paris* 342.10, pp. 723–726. MR: 2227747 (cit. on p. 304).

Winfried Bruns and Jürgen Herzog (1993). *Cohen-Macaulay rings*. Vol. 39. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, pp. xii+403. MR: 1251956 (cit. on pp. 302–304).

Sankar P. Dutta (1987). "On the canonical element conjecture". *Trans. Amer. Math. Soc.* 299.2, pp. 803–811. MR: 869233 (cit. on p. 304).

Lawrence Ein, Robert Lazarsfeld, and Karen E. Smith (2001). "Uniform bounds and symbolic powers on smooth varieties". *Invent. Math.* 144.2, pp. 241–252. MR: 1826369 (cit. on p. 305).

E. Graham Evans and Phillip Griffith (1981). "The syzygy problem". *Ann. of Math. (2)* 114.2, pp. 323–333. MR: 632842 (cit. on p. 304).

Gerd Faltings (2002). "Almost étale extensions". *Astérisque* 279. Cohomologies p-adiques et applications arithmétiques, II, pp. 185–270. MR: 1922831 (cit. on p. 297).

Ofer Gabber and Lorenzo Ramero (2003). *Almost ring theory*. Vol. 1800. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. vi+307. MR: 2004652 (cit. on pp. 296–298).

Raymond C. Heitmann (2001). "Extensions of plus closure". *J. Algebra* 238.2, pp. 801–826. MR: 1823785 (cit. on p. 306).

– (2002). "The direct summand conjecture in dimension three". *Ann. of Math. (2)* 156.2, pp. 695–712. MR: 1933722 (cit. on p. 295).

Raymond Heitmann and Linquan Ma (Mar. 2017a). "Big Cohen-Macaulay algebras and the vanishing conjecture for maps of Tor in mixed characteristic". arXiv: 1703.08281 (cit. on p. 305).

– (Aug. 2017b). "Extended plus closure in complete local rings". arXiv: 1708.05761 (cit. on p. 306).

M. Hochster (1973). "Contracted ideals from integral extensions of regular rings". *Nagoya Math. J.* 51, pp. 25–43. MR: 0349656 (cit. on pp. 295, 301).

Melvin Hochster (1983). "Canonical elements in local cohomology modules and the direct summand conjecture". *J. Algebra* 84.2, pp. 503–553. MR: 723406 (cit. on p. 296).

– (2002). "Big Cohen-Macaulay algebras in dimension three via Heitmann's theorem". *J. Algebra* 254.2, pp. 395–408. MR: 1933876 (cit. on p. 303).

– (2007). "Homological conjectures, old and new". *Illinois J. Math.* 51.1, pp. 151–169. MR: 2346192 (cit. on pp. 303, 304).

Melvin Hochster and Craig Huneke (1990). "Tight closure, invariant theory, and the Briançon-Skoda theorem". *J. Amer. Math. Soc.* 3.1, pp. 31–116. MR: 1017784 (cit. on p. 305).

– (1995). "Applications of the existence of big Cohen-Macaulay algebras". *Adv. Math.* 113.1, pp. 45–117. MR: 1332808 (cit. on pp. 302, 303, 305).

Takesi Kawasaki (2000). "On Macaulayfication of Noetherian schemes". *Trans. Amer. Math. Soc.* 352.6, pp. 2517–2552. MR: 1707481 (cit. on p. 302).

Kiran S. Kedlaya and Ruochuan Liu (2015). "Relative p-adic Hodge theory: foundations". *Astérisque* 371, p. 239. MR: 3379653 (cit. on p. 297).

Linquan Ma and Karl Schwede (May 2017). "Perfectoid multiplier/test ideals in regular rings and bounds on symbolic powers". arXiv: 1705.02300 (cit. on p. 305).

Takeo Ohi (1996). "Direct summand conjecture and descent for flatness". *Proc. Amer. Math. Soc.* 124.7, pp. 1967–1968. MR: 1317044 (cit. on p. 296).

Christian Peskine and Lucien Szpiro (1973). "Dimension projective finie et cohomologie locale". *Publications Mathématiques de l'Institut des Hautes Études Scientifiques* 42, pp. 323–395 (cit. on p. 304).

Michel Raynaud and Laurent Gruson (1971). "Critères de platitude et de projectivité. Techniques de "platification" d'un module". *Invent. Math.* 13, pp. 1–89. MR: 0308104 (cit. on p. 296).

Paul Roberts (1987). "Le théorème d'intersection". *C. R. Acad. Sci. Paris Sér. I Math.* 304.7, pp. 177–180. MR: 880574 (cit. on p. 304).

Peter Scholze (2012). "Perfectoid spaces". *Publ. Math. Inst. Hautes Études Sci.* 116, pp. 245–313. MR: 3090258 (cit. on pp. 296, 297, 300).
– (2015). "On torsion in the cohomology of locally symmetric varieties". *Ann. of Math. (2)* 182.3, pp. 945–1066. MR: 3418533 (cit. on p. 299).
Jean-Pierre Serre (1965). *Algèbre locale. Multiplicités*. Vol. 11. Cours au Collège de France, 1957–1958, rédigé par Pierre Gabriel. Seconde édition, 1965. Lecture Notes in Mathematics. Springer-Verlag, Berlin-New York, vii+188 pp. (not consecutively paged). MR: 0201468 (cit. on p. 303).
Kazuma Shimomoto (2016). "An application of the almost purity theorem to the homological conjectures". *J. Pure Appl. Algebra* 220.2, pp. 621–632. MR: 3399381 (cit. on p. 298).
– (June 2017). "Integral perfectoid big Cohen-Macaulay algebras via André's theorem". arXiv: 1706.06946 (cit. on p. 303).

Received 2017-11-30.

YVES ANDRÉ
INSTITUT DE MATHÉMATIQUES DE JUSSIEU, 4 PLACE JUSSIEU, 75005 PARIS FRANCE
yves.andre@imj-prg.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (309–338)

LA COURBE

Laurent Fargues

Résumé

On présente un résumé de nos travaux sur la courbe que nous avons introduite avec Jean-Marc Fontaine et ses applications en théorie de Hodge p-adique ainsi qu'au programme de Langlands.

Introduction

La courbe fondamentale en théorie de Hodge p-adique a été introduite par Jean-Marc Fontaine et l'auteur dans Fargues et Fontaine [2017]. On présente d'abord celle-ci, l'étude des fibrés vectoriels et de leurs modifications dans la Section 1. Dans la Section 2 on explique nos résultats sur l'étude des G-fibrés et leurs applications aux espaces de périodes p-adiques. Enfin, on conclut dans la Section 3 par des résultats récents sur les familles de fibrés et le lien conjectural avec la correspondance de Langlands locale. Ce texte est en quelque sorte un récit du parcours qui nous a mené des filtrations de Harder-Narasimhan des schémas en groupes finis et plats (Fargues [2010b]) à la conjecture de géométrisation de la correspondance de Langlands locale (Section 3).

1 Courbe et fibrés vectoriels

1.1 Fonctions holomorphes de la variable p. Soient E un corps local de corps résiduel le corps fini $\mathbb{F}_q$ et π une uniformisante de E. On a donc soit $[E : \mathbb{Q}_p] < +\infty$, soit $E = \mathbb{F}_q((\pi))$. Soit $F|\mathbb{F}_q$ un corps perfectoïde. On note $\varpi \in F$ une pseudo-uniformisante i.e. $0 < |\varpi| < 1$. Considérons l'anneau de Fontaine (Fontaine [1979], Fontaine [1994a])

$$\mathbf{A} = \begin{cases} W_{\mathcal{O}_E}(\mathcal{O}_F) \text{ si } E|\mathbb{Q}_p \\ \mathcal{O}_F[\![\pi]\!] \text{ si } E = \mathbb{F}_q((\pi)). \end{cases}$$

L'auteur a bénéficié du support des projets ANR-14-CE25 "PerCoLaTor" et ERC Advanced Grant 742608 "GeoLocLang".
MSC2010 : primary 11S31 ; secondary 11S37, 11G18, 14K10.

La théorie de la courbe schématique ne nécessite pas la théorie des espaces perfectoïdes qu'elle prédate ou des espaces adiques. Néanmoins il est plus naturel maintenant de l'exposer en introduisant ces objets. Considérons donc l'espace adique

$$Y = \mathrm{Spa}(\mathbf{A}, \mathbf{A}) \setminus V(\pi[\varpi]).$$

Si $E = \mathbb{F}_q(\!(\pi)\!)$ il s'agit d'un disque ouvert épointé de la variable π, $\mathbb{D}_F^* = \{0 < |\pi| < 1\} \subset \mathbb{A}_F^1$ qui est exactement celui qui apparaît dans les travaux de (Hartl et Pink [2004]). Néanmoins on ne le considère pas ici comme un espace adique sur F mais sur E,

$$\begin{array}{ccc} & Y = \mathbb{D}_F^* & \\ {}^{\text{loc. de type fini}}\swarrow & & \searrow^{\text{pas loc. de type fini}} \\ \mathrm{Spa}(F) & & \mathbb{D}_{\mathbb{F}_q}^* = \mathrm{Spa}(\mathbb{F}_q(\!(\pi)\!)). \end{array}$$

Il est préperfectoïde, si $E|\mathbb{Q}_p$ et E_∞ est le complété de l'extension engendrée par les points de torsion d'un groupe de Lubin-Tate, l'espace $Y \hat{\otimes}_E E_\infty$ est perfectoïde de basculé le même espace associé à $E_\infty^\flat$ i.e. $\mathbb{D}_F^{*,1/p^\infty}$ un disque perfectoïde épointé (Fargues [2015c] sec. 2.2).

L'algèbre de Fréchet $\mathcal{O}(Y)$ est obtenue par complétion des fonctions holomorphes sur Y, méromorphes le long des diviseurs (π) et $([\varpi])$,

$$\mathbf{A}[\tfrac{1}{\pi}, \tfrac{1}{[\varpi]}],$$

relativement aux normes de Gauss $|.|_\rho$ pour des rayons $\rho \in]0,1[$

$$\Big| \sum_{n \gg -\infty} [x_n]\pi^n \Big|_\rho = \sup_n |x_n|\rho^n.$$

Un élément $\xi = \sum_n [x_n]\pi^n \in \mathbf{A}$ est dit primitif de degré $d > 0$ si $x_0 \neq 0$, $x_0, \dots, x_{d-1} \in \mathfrak{m}_F$ et $x_d \in \mathcal{O}_F^\times$. Le produit d'un élément primitif de degré d par un de degré d' est primitif de degré $d + d'$ et on a donc une bonne notion d'élément primitif irréductible. Deux tels éléments sont équivalents s'ils sont multiples par une unité dans $\mathbf{A}^\times$. Le résultat suivant est un un résultat clef pour toute la suite (on renvoie également aux articles de revue Fargues et Fontaine [2012], Fargues et Fontaine [2011] et Fargues et Fontaine [2014]).

Théorème 1.1 (Fargues et Fontaine [2017] théo. 2.4, coro. 2.2.23, théo. 3.1.11)**.**

1. *Si F est algébriquement clos alors tout élément primitif dans $\mathbf{A}$ irréductible est de degré 1, équivalent à $\pi - [a]$ avec $0 < |a| < 1$. En d'autres termes, tout ξ primitif a une factorisation (de Weierstrass)*

$$\xi = u(\pi - [a_1]) \dots (\pi - [a_d]),\ u \in \mathbf{A}^\times, 0 < |a_i| < 1.$$

2. *Pour ξ primitif irréductible dans* $\mathbf{A}$*,* $\mathbf{A}[\frac{1}{\pi}]/(\xi) = \mathcal{O}(Y)/(\xi)$ *est un corps perfectoïde de basculé une extension de degré* $\deg(\xi)$ *de* F *via* $x \mapsto [x]$ *mod* ξ*. Cela induit une équivalence*

$$\{\text{éléments primitifs irréductibles de degré 1}\}/\sim \xrightarrow{\sim} \text{débasculements de } F \text{ sur } E.$$

Par définition les points classiques de Y, $|Y|^{cl}$, sont les zéros des éléments primitifs de $\mathbf{A}$. Les normes de Gauss $|.|_\rho$ sont multiplicatives et *on dispose donc d'une bonne notion de polygone de Newton pour les fonctions holomorphes* f *sur* Y ainsi que sur ses couronnes, ceci par application d'une transformée de Legendre inverse appliquée à la fonction concave $]0,+\infty[\ni r \mapsto -\log|f|_{q^{-r}}$. Par « bonne notion de polygone de Newton » on entend typiquement le fait que le polygone d'un produit est le concaténé des polygones (Fargues et Fontaine [ibid., sec. 1], Fargues et Fontaine [2014, sec. 1]). Par exemple, le polygone de Newton de $\sum_{n\gg 0}[x_n]\pi^n \in \mathbf{A}[\frac{1}{\pi}, \frac{1}{[\varpi]}]$ est l'enveloppe convexe décroissante des points $(n, v(x_n))_{n\in\mathbb{Z}}$. Si $\xi = \sum_n [x_n]\pi^n$ est primitif de degré d on définit sa valuation (la « distance à l'origine $\pi = 0$ dans le disque épointé Y ») comme étant $\frac{1}{d}v(x_0)$.

Théorème 1.2 (Fargues et Fontaine [2017, théo. 3.4.4])**.**

1. *Pour* $f \in \mathcal{O}(Y)$ *non nul les pentes de son polygone de Newton sont les valuations des zéros de* f *dans* $|Y|^{cl}$*,* $f(y) = 0$ *avec* $y \in |Y|^{cl}$*, comptées avec multiplicité* $\mathrm{ord}_y(f)\deg(y)$ *où la valuation* ord_y *est celle de l'anneau de périodes de Fontaine* $\widehat{\mathcal{O}}_{Y,y} = B_{dR}^+(k(y))$ *associé au corps perfectoïde* $k(y)$*.*
2. *L'algèbre de Banach des fonctions holomorphes sur une couronne compacte* $\{\rho_1 \leq |\pi| \leq \rho_2\}$ *de* Y *est un anneau principal de spectre maximal les éléments de* $|Y|^{cl}$ *dans cette couronne (à une unité près tout élément irréductible est primitif irréductible dans* $\mathbf{A}$*).*

Concrètement, si F est algébriquement clos et si λ est une pente du polygone de f alors $f = (\pi - [a])g$ avec $v(a) = \lambda$ et la multiplicité de λ dans le polygone de g est une de moins. L'analyse p-adique d'une fonction d'une variable développée dans Lazard [1962] s'étend à ce cadre. On peut typiquement former des produits de Weierstrass et vérifier ainsi que tout $f \in \mathcal{O}(Y)$ se met sous la forme

$$\prod_{n\geq 0}\left(1 - \frac{[x_n]}{\pi}\right) \times g$$

avec $x_n \to 0$, g sans zéros au voisinage de $\pi = 0$ et donc méromorphe en $\pi = 0$, de la forme $\sum_{n\gg 0}[x_n]\pi^n$ avec $x_n \in F$ et pour tout $\rho \in]0,1[$, $\lim_{n\to+\infty}|x_n|\rho^n = 0$ (Fargues et Fontaine [2017] sec. 1.2). Cela permet de rendre concret les éléments de $\mathcal{O}(Y)$ puisqu'*en général, sauf si bien sûr si* $E = \mathbb{F}_q((\pi))$*, un tel élément n'admet pas de développement unique en série de Laurent de la forme* $\sum_{n\in\mathbb{Z}}[x_n]\pi^n$.

1.2 La courbe schématique. L'espace Y est muni d'un Frobenius φ induit par le Frobenius $x \mapsto x^q$ de F. *Les théorèmes de factorisation précédents permettent d'analyser les périodes p-adiques*. Typiquement un élément de $\mathcal{O}(Y)^{\varphi=\pi^d}$ a un polygone de Newton satisfaisant une équation fonctionnelle qui se résout facilement. On peut alors lui appliquer les résultats de factorisation précédent. Utilisant ces résultats on obtient le théorème clef suivant.

Théorème 1.3 (Fargues et Fontaine [2017, théo. 6.2.1])**.** *Si le corps F est algébriquement clos l'algèbre de périodes $P = \bigoplus_{d\geq 0} \mathcal{O}(Y)^{\varphi=\pi^d}$ est graduée factorielle i.e. le monoïde abélien $\coprod_{d\geq 0} P_d \setminus \{0\}/E^\times$ est libre de base $P_1 \setminus \{0\}/E^\times$.*

Concrètement, si $f \in P_d \setminus \{0\}$, $f = f_1 \dots f_d$ avec $f_i \in P_1$ et une telle écriture est unique à des scalaires dans $E^\times$ près. Ce théorème (qui se formule et se démontre sans espaces adiques) s'est récemment retrouvé de nouveau au coeur de résultats plus complexes à priori, cf. proposition 3.2. *Il est au coeur même de la courbe.*

De la même façon, en utilisant ces résultats d'analyse concernant les fonctions de $\mathcal{O}(Y)$ et leurs zéros, on donne une démonstration rapide de la suite exacte fondamentale de la théorie de Hodge p-adique dans (Fargues et Fontaine [ibid., théo. 6.4.1]). En combinant ces divers résultats on obtient la construction de la courbe schématique.

Théorème 1.4 (Fargues et Fontaine [ibid., théo. 6.5.2, 7.3.3])**.** *Le schéma $X = Proj(P)$ est noethérien régulier de dimension 1, i.e. de Dedekind. De plus*

1. *Il y a un morphisme d'espaces annelés $Y \to X$ qui induit une bijection $|Y|^{cl}/\varphi^{\mathbb{Z}} \to |X|$ (points fermés de X) tel que si $y \mapsto x$ alors $k(y) = k(x)$ et $B_{dR}^+(k(y)) = \widehat{\mathcal{O}}_{Y,y} = \widehat{\mathcal{O}}_{X,x}$.*
2. *La courbe est « complète » au sens où si pour $x \in |X|$ on pose $\deg(x) := [k(x)^\flat : F]$, alors pour tout $f \in E(X)^\times$, $\deg(\div f) = 0$.*
3. *Si F est algébriquement clos alors tout point est de degré 1, $P_1 \setminus \{0\}/E^\times \xrightarrow{\sim} |X|$ via $t \mapsto V^+(t)$. De plus $\mathcal{O}_X(D^+(t)) = \mathcal{O}(Y)[\frac{1}{t}]^{\varphi=Id} = B_{cris}(C_t)^{\varphi=Id}$ est un anneau principal où C_t est le corps résiduel en $V^+(t)$, $C_t^\flat = F$.*

1.3 La courbe adique. Comme on l'a dit précédemment l'espace Y est adique préperfectoïde. On définit alors la courbe adique comme étant

$$X^{ad} = Y/\varphi^{\mathbb{Z}}.$$

C'est un espace adique quasicompact partiellement propre (mais pas de type fini) sur $\mathrm{Spa}(E)$. Il y a un morphisme d'espaces annelés $X^{ad} \to X$ qui identifie les points classiques de X^{ad} et les points fermés de X ainsi que les complétés des anneaux locaux correspondants. Bien que non localement de type fini on sait que Y *et* X^{ad} *sont localement* *noethériens* (Fargues [2015c] conjecture 1 démontrée par Kedlaya [2016a]). L'un des intérêts de ce résultat est que du coup Y et X^{ad} rentrent dans le cadre de la théorie « classique » développée par Huber [1996]. Ainsi, par exemple, si Z est un E-espace rigide de Tate, i.e. un E-espace adique localement de type fini, on peut former l'espace adique $X^{ad} \times Z$ et regarder les faisceaux cohérents dessus lorsque Z varie (cf. Kedlaya, Pottharst et Xiao [2014] pour une exemple d'une version non perfectoïde de cela).

1.4 Fibrés vectoriels. Bien que l'algèbre graduée de périodes P utilisée pour définir X dépende du choix de π, X n'en dépend canoniquement pas. Ceci dit, le choix de π définit un fibré « très ample » $\mathcal{O}(1) = \widetilde{P[1]}$ sur X tel que

$$P = \bigoplus_{d \geq 0} H^0(X, \mathcal{O}(d)).$$

On suppose maintenant que F contient une clôture algébrique $\overline{\mathbb{F}}_q$ de $\mathbb{F}_q$. On note E_n l'extension non ramifiée de degré n de E et $\breve{E} = \widehat{E^{nr}}$ muni de son Frobenius σ. On a une identification (l'indice E signifie la courbe associée au choix de E)

$$X_E \otimes_E E_n = X_{E_n}$$

et le revêtement cyclique $X^{ad}_{E_n} \to X^{ad}_E$ s'identifie à

$$Y/\varphi^{n\mathbb{Z}} \longrightarrow Y/\varphi^{\mathbb{Z}}.$$

Ainsi,

trivialisation du revêtement $Y \to Y/\varphi^{\mathbb{Z}}$

X^{ad} — ext. des scalaires non ramifiées $\nearrow$

X^{ad} — ext. des scalaires ramifiées $\searrow$

monde perfectoïde.

Pour tout $\lambda = \frac{d}{h} \in \mathbb{Q}$, $(d,h) = 1$, on définit $\mathcal{O}_X(\lambda)$ comme étant le poussé en avant par le revêtement étale fini $X_{E_h} \to X_E$ de $\mathcal{O}_{X_{E_h}}(d)$. Un des points fondamentaux de la théorie est maintenant le suivant : puisque la courbe est complète (théorème 1.4) on peut définir naturellement le degré d'un fibré en droites et *on dispose en particulier du formalisme des filtrations de Harder-Narasimhan* (Fargues et Fontaine [2017] sec. 5.5). Ainsi *le fibré $\mathcal{O}(\lambda)$ est stable de pente λ.*

Théorème 1.5 (Fargues et Fontaine [ibid.] théo. 8.2.10)**.** *Supposons F algébriquement clos.*

1. *Pour tout $\lambda \in \mathbb{Q}$, un fibré sur X est semi-stable de pente λ si et seulement si il est isomorphe à une somme directe de $\mathcal{O}(\lambda)$.*
2. *La filtration de Harder-Narasimhan d'un fibré sur X est scindée.*
3. *L'application $(\lambda_i)_i \mapsto \oplus_i \mathcal{O}(\lambda_i)$ induit une bijection entre suites décroissantes de nombres rationnels et classes d'isomorphisme de fibrés sur X.*

Ce résultat ressemble fortement à une généralisation du théorème de Grothendieck concernant la droite projective. Néanmoins sa preuve est beaucoup plus complexe car contrairement à $\mathbb{P}^1$, $H^1(X, \mathcal{O}(-1)) \neq 0$: si $V^+(t) = \infty$ avec $t \in P_1 \setminus \{0\}$ alors *l'anneau principal* $\mathcal{O}(Y)[1/t]^{\varphi=Id}$ *muni du stathme* $-\mathrm{ord}_\infty$ *n'est pas euclidien !* La démonstration de ce résultat utilise des résultats fins de théorie de Hodge p-adique concernant les périodes de groupes p-divisibles et leurs liens avec les modifications de fibrés (cf. Section 1.7.2).

Lorsque le corps F n'est plus algébriquement clos des *phénomènes monodromiques apparaissent.* On vérifie typiquement que pour $t \in P_1 \setminus \{0\}$

$$\mathrm{Cl}(\mathcal{O}(Y)[1/t]^{\varphi=Id}) = \mathrm{Pic}^0(X) = \mathrm{Hom}(\mathrm{Gal}(\overline{F}|F), E^\times).$$

On a plus généralement le résultat suivant.

Théorème 1.6 (Fargues et Fontaine [2017] théo. 9.3.1, du type Narasimhan Narasimhan et Seshadri [1965])**.** *Via le foncteur $\mathcal{E} \mapsto H^0(X_{\widehat{\overline{F}}}, \mathcal{E}_{|X_{\widehat{\overline{F}}}})$, la catégorie des fibrés semi-stables de pente 0 sur X_F est équivalente à celle des représentations p-adiques de* $\mathrm{Gal}(\overline{F}|F)$ *à coefficients dans E.*

Enfin, venons en au théorème GAGA.

Théorème 1.7 (GAGA, Fargues [2015c] théo. 3.5, Kedlaya et Liu [2015] théo. 8.5.5)**.** *L'image réciproque par le morphisme d'espaces annelés $X^{ad} \to X$ induit une équivalence entre faisceaux cohérents sur X et X^{ad}.*

La première démonstration de ce résultat (Fargues [2015c]) consistait à comparer le théorème de classification 1.4 avec celui de Kedlaya [2004] lorsque $E|\mathbb{Q}_p$ ou bien le théorème d'Hartl Pink en égales caractéristiques (Hartl et Pink [2004]). Plus précisément, les fibrés vectoriels sur X^{ad} s'identifient aux fibrés φ-équivariants sur Y, qui eux-même s'identifient aux germes de fibrés vectoriels φ-équivariants au voisinage de $\pi = 0$ sur Y (si U est un tel voisinage alors $Y = \cup_{n \geq 0} \varphi^n(U)$). *L'anneau des germes de fonctions holomorphes au voisinage de $\pi = 0$ s'identifie à l'anneau de Robba* associé à F, et donc les fibrés vectoriels sur X^{ad} s'identifient aux φ-modules sur cet anneau de Robba.

Kedlaya et Liu [2015] ont donné une autre démonstration reposant sur leur preuve du fait que $\mathcal{O}(1)$ *est ample sur X^{ad}* : pour tout fibré $\mathcal{E}$ sur X^{ad}, pour $d \gg 0$, $\mathcal{E}(d)$ est engendré par ses sections globales (bien sûr ce résultat est tautologique sur la courbe schématique, c'est une des raisons pour laquelle elle est plus simple à manipuler, ce résultat est intégré dans sa construction). Il s'agit de construire suffisamment de sections de $\mathcal{E}(d)$, $d \gg 0$, par un processus itératif convergeant (KEDLAYA et LIU [ibid.] sec. 6.2), les méthodes usuelles de séries de Poincaré (CARTAN [1958]) sur Y ne semblant pas fonctionner (de plus, si F est algébriquement clos, on ne sait pas à l'avance que le tiré en arrière à Y d'un fibré sur X^{ad} est trivial, i.e. qu'un tel fibré est donné par un facteur d'automorphie, alors que c'est une conséquence du théorème de classification).

Une façon plus intrinsèque d'exposer le théorème de classification des fibrés est d'utiliser les isocristaux tels qu'ils apparaissent dans le théorème de Dieudonné-Manin ; $\varphi\text{−Mod}_{\breve{E}}$ qui est la catégorie des couples (D, φ) où D est un $\breve{E}$–espace vectoriel de dimension finie et φ un automorphisme σ-linéaire. L'espace Y vit au dessus de $\text{Spa}(\breve{E})$. Dès lors on peut construire pour un tel isocristal

$$\mathcal{E}(D, \varphi) = Y \times_{\varphi^{\mathbb{Z}}} D \longrightarrow Y/\varphi^{\mathbb{Z}} = X^{ad}$$

qui est un fibré trivialisé de fibre D sur Y ayant pour facteur d'automorphie φ agissant sur D. Via GAGA cela correspond au fibré associé au P-module gradué $\oplus_{d \geq 0}(D \otimes \mathcal{O}(Y))^{\varphi \otimes \varphi = \pi^d}$. Dès lors le théorème 1.5 s'énonce en disant que

$$\mathcal{E}(-) : \varphi\text{−Mod}_{\breve{E}} \longrightarrow \text{Fibrés sur } X$$

est essentiellement surjectif.

1.5 Simple connexité géométrique de la courbe.

Du théorème de classification des fibrés on peut déduire le résultat suivant.

Théorème 1.8 (FARGUES et FONTAINE [2017] théo. 8.6.1)**.** *Si F est algébriquement clos le schéma $X_{\overline{E}}$ est simplement connexe : tout revêtement étale fini est scindé.*

En d'autres termes $\pi_1(X) = \text{Gal}(\overline{E}|E)$. Plus généralement, pour F quelconque, $\pi_1(X) = \text{Gal}(\overline{F}|F) \times \text{Gal}(\overline{E}|E)$ (FARGUES et FONTAINE [ibid.] théo. 9.5.1). On dispose du même résultat, soit par la même méthode soit par GAGA, sur la courbe adique : les revêtements étales finis de $\text{Spa}(E)$ correspondent à ceux de X^{ad}. Ce résultat a d'importantes applications. Il se réinterprète de façon formelle en disant que si F est algébriquement clos alors le diamant au sens de SCHOLZE [p. d., 2014] (cf. FARGUES [2017] par exemple pour ce diamant en particulier et la Section 3.3)

$$\text{Div}^1_F = \text{Spa}(F) \times \text{Spa}(E)^{\diamond}/\varphi^{\mathbb{Z}}_{E^{\diamond}}$$

a pour π_1 le groupe de Galois $\mathrm{Gal}(\overline{E}|E)$ ($X_F^{ad,\diamond}$ et Div_F^1 ont même catégorie de revêtements étales). On a

$$\mathrm{Spa}(F) \times \mathrm{Spa}(E)^\diamond = \mathbb{D}_F^{*,1/p^\infty}/\mathcal{O}_E^\times$$

un disque perfectoïde épointé divisé par l'action de Lubin-Tate de $\mathcal{O}_E^\times$ (Fargues [2015c] sec.2 et Scholze [2014]). Dès lors, si $F = \mathbb{C}_p^\flat$ on trouve comme corollaire du théorème 1.8 que *le groupe de Galois* $\mathrm{Gal}(\overline{E}|E)$ *classifie les revêtements étales finis* $\mathcal{O}_E^\times$-*équivariants de* $\mathbb{D}_{\mathbb{C}_p}^*/\varphi^{\mathbb{Z}}$ (Weinstein [2017]). Ce résultat a motivé l'article Kucharczyk et Scholze [2016] de Kucharczyk et Scholze qui est la recherche d'un résultat analogue sur les corps de nombres.

Le théorème 1.8 est également à la base de la preuve du *lemme de Drinfeld-Scholze* (Scholze [2014]) qui affirme que « $\pi_1\big((\mathrm{Div}^1)^d\big) = \pi_1(\mathrm{Div}^1)^d$ » (cf. Fargues [2017] sec. 4 pour une signification précise de cet énoncé en termes de Div^1).

1.6 Fibrés et espaces de Banach-Colmez. La découverte de la courbe remonte à l'étude par Fontaine et l'auteur des résultats de Colmez sur les espaces de Banach de dimension finie (Colmez [2002], Colmez [2008]). C'est en définissant des filtrations de Harder-Narasimhan sur ceux-ci et en tentant de les classifier que nous sommes tombés sur la courbe. Maintenant, le lien entre fibrés vectoriels sur la courbe et espaces de Banach-Colmez est complètement clarifié par le théorème suivant.

Théorème 1.9 (Le Bras [2017])**.** *Supposons que* $F = C^\flat$ *avec* C *algébriquement clos.*

1. *La catégorie des espaces de Banach-Colmez est la plus petite sous-catégorie abélienne de la catégorie des faisceaux pro-étales de* E*-espaces vectoriels sur* $\mathrm{Spa}(C)$ *contenant* $\mathbb{G}_a$, $\underline{E}$ *et stable par extensions.*
2. *Le foncteur cohomologie relative induit une équivalence entre le coeur dans* $\mathbb{D}^b_{coh}(\mathcal{O}_X)$ *de la* t*-structure*
   - *dont la partie* ≤ 0 *est formée des complexes* $\mathcal{E}^\bullet$ *satisfaisant* $\mathcal{H}^i(\mathcal{E}^\bullet) = 0$ *si* $i > 0$ *et* $\mathcal{H}^0(\mathcal{E}^\bullet)$ *est à pentes de HN* ≥ 0,
   - *la partie* ≥ 0 *des complexes satisfaisant* $\mathcal{H}^i(\mathcal{E}^\bullet) = 0$ *si* $i < -1$ *et* $\mathcal{H}^{-1}(\mathcal{E}^\bullet)$ *est à pentes de HN* < 0,

   et la catégorie des espaces de Banach-Colmez.
3. *L'équivalence précédente s'étend en une équivalence dérivée.*

Ici par cohomologie relative on entend la chose suivante. On peut définir un morphisme de (gros) topos pro-étales

$$\tau : (X_{C^\flat}^{ad})^\sim_{\text{pro-ét}} \longrightarrow \mathrm{Spa}(C^\flat)^\sim_{\text{pro-ét}} = \mathrm{Spa}(C)^\sim_{\text{pro-ét}}$$

grâce à la fonctorialité de la courbe adique : $S \mapsto X_S^{ad}$ (cf. Section 3.1) où S est un espace perfectoïde sur $\mathrm{Spa}(\mathbb{F}_q)$. Dès lors, si $\mathcal{E}^\bullet$ est un complexe de faisceaux cohérents sur $X_{C^\flat}^{ad}$ on peut définir $R\tau_*\mathcal{E}^\bullet$.

Certains de ces faisceaux de cohomologie relative, ceux associés aux fibrés vectoriels à pentes dans $[0,1]$, sont représentables par des C-espaces perfectoïdes du type $\varprojlim_{\times p} \mathcal{G}^{rig}$ où $\mathcal{G}$ est un groupe formel p-divisible sur $\mathcal{O}_C$ (cf. sec. 4 de FARGUES et FONTAINE [2017], c'est ce que Scholze et Weinstein ont appelé plus tard le revêtement universel d'un groupe p-divisible (SCHOLZE et WEINSTEIN [2013]). Néanmoins en général ils ne sont représentables que par des diamants (et ces espaces de Banach-Colmez ont probablement été une forte source d'inspiration pour l'introduction par Scholze de la théorie des diamants).

Via le théorème de classification des fibrés 1.5 cela fournit une classification des espaces de Banach-Colmez. Remarquons que la t-structure intervenant dans ce dernier théorème est analogue à celle utilisée par Bridgeland (BRIDGELAND [2006]) qui a été une inspiration pour ce résultat.

Dans ce cadre là, Fontaine a récemment fait le lien entre le coeur de la t-structure analogue sur les fibrés Galois équivariants et sa théorie des presque $\mathbb{C}_p$-représentations (FONTAINE [2003]). Cela permet de boucler la boucle puisque cette théorie a été une forte inspiration pour la théorie des espaces de Banach-Colmez (FONTAINE [ibid.] sec. 4.1 par exemple).

Enfin, les résultats de BERGER [2008a] clarifient complètement le lien entre fibrés Galois équivariants et la théorie « classique » des (φ, Γ)-modules : si K est de valuation discrète à corps résiduel parfait, $C = \widehat{\overline{K}}$ et $F = C^\flat$ alors la catégorie des fibrés $\mathrm{Gal}(\overline{K}|K)$-équivariants sur la courbe est équivalente à celle des (φ, Γ)-modules.

1.7 Modifications de fibrés et théorie de Hodge p-adique.

1.7.1 Faiblement admissible implique admissible. La première application que nous avons donnée de la courbe avec Fontaine dans FARGUES et FONTAINE [2017], sec. 10, est une nouvelle preuve du théorème « faiblement admissible équivalent à admissible » de COLMEZ et FONTAINE [2000], BERGER [2008b]). C'était la seconde fois, après la preuve du théorème de classification des fibrés, que sont apparues naturellement les modifications de fibrés qui sont devenues essentielles dans la suite (dans l'exposé FARGUES [2010a], p.53, on dit appliquer une correspondance de Hecke lorsqu'on modifie un fibré).

Plus précisément, prenons $E = \mathbb{Q}_p$ et soit $K|\breve{\mathbb{Q}}_p$ une extension de degré fini. On note $C = \widehat{\overline{K}}$ et $\Gamma = \mathrm{Gal}(\overline{K}|K)$. La courbe associée à $C^\flat$ est munie d'une action de Γ et d'un point fermé ∞ de corps résiduel C invariant sous l'action de Γ. Cette action « arithmétique » de Γ rigidifie complètement la situation. On a en effet les propriétés suivantes :

1. ∞ est le seul point fermé dont la Γ-orbite soit finie et donc la catégorie des fibrés Γ-équivariants sur $X \setminus \{\infty\}$ est abélienne (FARGUES et FONTAINE [2017] 10.1.1, 10.1.3)
2. si $(D,\varphi) \in \varphi\text{−Mod}_{\breve{\mathbb{Q}}_p}$ alors les sous-fibrés Γ-invariants de $\mathcal{E}(D,\varphi)_{|X\setminus\{\infty\}}$ sont en bijection avec les sous-isocristaux de (D,φ) (FARGUES et FONTAINE [ibid.] théo. 10.2.14).

Soit donc $(D,\varphi) \in \varphi\text{−Mod}_{\breve{\mathbb{Q}}_p}$ un isocristal. On s'intéresse aux modifications Γ-équivariantes de $\mathcal{E}(D,\varphi)$ en ∞. Ce fibré est trivialisé en ∞ et $\widehat{\mathcal{E}(D,\varphi)}_\infty = D \otimes B_{dR}^+ = D_K \otimes_K B_{dR}^+$. On dispose maintenant de l'énoncé suivant : pour W un K-espace vectoriel de dimension finie il y a une bijection

$$\begin{aligned}\text{Filtrations de W} &\xrightarrow{\sim} \text{réseaux } \Gamma\text{-invariants dans } W \otimes B_{dR}\\ \text{Fil}^\bullet W &\longmapsto \text{Fil}^0(W \otimes B_{dR}).\end{aligned}$$

Du point de vue « de la grassmanienne affine » (SCHOLZE [2014] où Scholze montre que l'on peut mettre une structure de ind-diamant sur ces objets), si μ est un cocaractère diagonal dominant à valeurs dans GL_n et

$$\text{Gr}_\mu = \text{GL}_n(B_{dR}^+)\mu(t)\text{GL}_n(B_{dR}^+)/GL_n(B_{dR}^+)$$

est une cellule de Schubert affine (comme ensemble), il y a une application de Bialynicki-Birula (CARAIANI et SCHOLZE [2017, prop. 3.4.3] pour la version en familles pour n'importe quel groupe)

$$\text{Gr}_\mu \longrightarrow \text{GL}_n(C)/P_\mu(C).$$

Le résultat précédent dit que cela induit une bijection

$$\text{Gr}_\mu^\Gamma \xrightarrow{\sim} \text{GL}_n(K)/P_\mu(K).$$

Il s'agit là d'un résultat analogue au résultat « classique » sur $\mathbb{C}$ qui dit que pour l'action de $\mathbb{G}_m$ sur la cellule affine Gr_μ, avec ici $\text{Gr}(\mathbb{C}) = \text{GL}_n\big(\mathbb{C}(\!(t)\!)\big)/\text{GL}_n\big(\mathbb{C}[\![t]\!]\big)$, induite par $\lambda.t = \lambda t$ avec $\lambda \in \mathbb{G}_m$, alors

$$\text{Gr}_\mu^{\mathbb{G}_m} \xrightarrow{\sim} G/P_\mu.$$

Du point de vue des espaces de lacets cette action algébrique de $\mathbb{G}_m$ correspond à l'action de $U(1)$ de rotation des lacets. *Il y a donc une forte analogie entre $U(1)$ et Γ.* Par exemple, pour $\sigma \in \Gamma$, $\sigma(t) = \chi_{cyc}(\sigma)t$ où ici t est le $2i\pi$ p-adique de Fontaine.

Il résulte de cela que les modifications Γ-équivariantes de $\mathcal{E}(D,\varphi)$ sont en bijection avec les filtrations de D_K. Partant donc d'un φ-module filtré $(D,\varphi,\text{Fil}^\bullet D_K)$ au sens de Fontaine (Fontaine [1994b]) on en déduit par modification un fibré Γ-équivariant $\mathcal{E}(D,\varphi,\mathrm{Fil}^\bullet D_K)$. Sa filtration de Harder-Narasimhan est Γ-invariante et provient donc d'une filtration de (D,φ). Partant de là « le théorème faiblement admissible » implique admissible est une conséquence facile du théorème de classification des fibrés qui implique que $\dim_E H^0(X,\mathcal{F}) = \mathrm{rg}(\mathcal{F})$ ssi $\mathcal{F}$ est semi-stable de pente 0.

On donne également dans Fargues et Fontaine [2017], sec. 10.6, une preuve du théorème de la monodromie p-adique (de Rham implique potentiellement semi-stable (Colmez [2003])) en utilisant la courbe.

1.7.2 Modifications de fibrés et groupes p-divisibles. La preuve du théorème de classification des fibrés utilise de façon fondamentale deux résultats sur les modifications de fibrés qui se déduisent de résultats sur les périodes de groupes p-divisibles. C'est au cours de cette preuve que l'auteur a commencé à s'intéresser aux liens entre groupes p-divisibles et modifications de fibrés. Plus précisément, supposons que $E = \mathbb{Q}_p$ et $F = C^\flat$ avec $C|\mathbb{Q}_p$ algébriquement clos. Si $\mathrm{BT}_{\mathcal{O}_C}$ désigne les groupes de Barsotti-Tate sur $\mathcal{O}_C$ il y a un foncteur

$$\mathfrak{M} : \mathrm{BT}_{\mathcal{O}_C} \otimes \mathbb{Q} \longrightarrow \left\{ \begin{array}{c} \text{modifications minuscules de fibrés en } \infty \\ \mathcal{F} \hookrightarrow \mathcal{E} \text{ avec } \mathcal{F} \text{ semi-stable de pente 0 i.e. trivial} \end{array} \right\}.$$

Ce foncteur admet les deux descriptions suivantes, si $\mathcal{G} \mapsto [\mathcal{F} \hookrightarrow \mathcal{E}]$,

— **[Périodes de de Rham]** On a $\mathcal{F} = V_p(\mathcal{G}) \otimes \mathcal{O}_X$ et $\mathcal{E} = \mathcal{E}(D, p^{-1}\varphi)$ où (D,φ) est le module de Dieudonné covariant de $\mathcal{G}_{k_C}$. Le morphisme $\mathcal{F} \hookrightarrow \mathcal{E}$ est alors donné par une application linéaire

$$V_p(\mathcal{G}) \to H^0(\mathcal{E}(D, p^{-1}\varphi)) = (D \otimes \mathcal{O}(Y))^{\varphi = p}$$

qui est une application de périodes cristallines de Fontaine. Le faisceau $\mathrm{coker}(\mathcal{F} \to \mathcal{E})$ est alors le faisceau gratte-ciel $\mathrm{Lie}(\mathcal{G})[\frac{1}{p}]$ en ∞.

— **[Périodes de Hodge-Tate]** Il y a un morphisme de Hodge-Tate $\alpha_{\mathcal{G}^D} : V_p(\mathcal{G}^D) \to \omega_{\mathcal{G}}[\frac{1}{p}]$ (Fontaine [1981], Fargues [2008], Fargues [2011]) qui induit une surjection $V_p(\mathcal{G}^D) \otimes \mathcal{O}_X \twoheadrightarrow i_{\infty *}\omega_{\mathcal{G}}[\frac{1}{p}]$ et donc une modification. La modification duale de cette modification est $[\mathcal{F} \hookrightarrow \mathcal{E}]$.

Ces deux types de périodes ont été unifiées plus tard via l'introduction des φ-modules sur A_{inf} (cf. Section 1.7.3). Le théorème de classification des fibrés utilise les deux résultats suivants :

1. *(Lafaille/Gross-Hopkins)* Tout élément de $\mathbb{P}^{n-1}(C)$ est la période de de Rham d'un groupe formel de hauteur n et de dimension 1

2. *(Drinfeld)* Tout élément de $\mathbb{P}^{n-1}(C) \setminus \bigcup_{H \in \check{\mathbb{P}}^{n-1}(\mathbb{Q}_p)} H(C)$ est la période de de Rham d'un $\mathcal{O}_D$-module formel spécial et donc (principe des tours jumelles, cf. Section 2.3) la période de Hodge-Tate d'un groupe formel de hauteur n et de dimension 1.

Via le foncteur $\mathfrak{M}$ le point (1) implique que toute modification de degré 1 de $\mathcal{O}(\frac{1}{n})$ est isomorphe à $\mathcal{O}^n$. Le point (2) implique que tout modification de degré -1 de $\mathcal{O}^n$ est isomorphe à $\mathcal{O}^{n-r} \oplus \mathcal{O}(\frac{1}{r})$, $1 \leq r \leq n$ (cf. sec. 8.3 de Fargues et Fontaine [2017]). Réciproquement, le théorème de classification fournit de nouveaux résultats concernant les espaces de périodes p-adiques (cf. Section 2.3). Ainsi, *la courbe est une machine qui recycle les deux résultats précédents pour en produire de nouveaux.*

La catégorie de modifications précédente dans le but de $\mathfrak{M}$ s'identifie (du point de vue des périodes de Hodge-Tate) à celle des couples (V, W) où : V est un $\mathbb{Q}_p$-espace vectoriel de dimension finie et $W \subset V_C$ est un sous-espace. On introduit dans Fargues [2013] la notion de *groupes rigides analytiques de type p-divisible.* Il s'agit de groupes rigides analytiques abéliens G sur C tels que

— $\times p : G \to G$ est surjectif de noyau fini
— $\times p : G \to G$ est topologiquement nilpotent i.e. « G est topologiquement de p^∞-torsion ».

On a alors le théorème suivant.

Théorème 1.10 (Fargues [ibid.]). *1. Le foncteur $\mathcal{G} \mapsto \mathcal{G}^{rig}$ de la catégorie des groupes formels p-divisibles sur $\mathcal{O}_C$ vers les groupes rigides analytiques de type p-divisible est pleinement fidèle, d'image essentielle les groupes G tels que $G \simeq \mathring{\mathbb{B}}^d_C$ comme espaces rigide pour un entier d.*

2. La catégorie des groupes rigides analytiques p-divisibles s'identifie à celle des triplets (Λ, W, u) où Λ est un $\mathbb{Z}_p$-module libre de rang fini, W un C-espace vectoriel de dimension finie et $u : W \to \Lambda_C$ est C-linéaire.

Dans le point (2), au triplet (Λ, W, u) on associe le groupe rigide analytique $(\log \otimes Id)^{-1}(W \otimes \mathbb{G}_a)$ dans la suite exacte

$$0 \longrightarrow \big(\Lambda[\tfrac{1}{p}]/\Lambda\big)(1) \longrightarrow \widehat{\mathbb{G}}_m^{rig} \otimes_{\mathbb{Z}_p} \Lambda \xrightarrow{\log \otimes Id} \mathbb{G}_a \otimes \Lambda_C \longrightarrow 0.$$

Les groupes G associés aux triplets (Λ, W, u) avec u injectif correspondent aux groupes G « hyperboliques » i.e. tels que tout morphisme rigide analytique $\mathbb{A}^1 \to G$ soit constant. Les groupes rigides analytiques de type p-divisible hyperboliques à isogénie près sont donc équivalents aux modifications minuscules dans le but du foncteur $\mathfrak{M}$. Ce même foncteur $\mathfrak{M}$ est donc de plus pleinement fidèle d'après le point (1) du théorème précédent. Scholze et Weinstein on complété ce résultat en montrant qu'il est en fait essentiellement surjectif.

Théorème 1.11 (Scholze et Weinstein [2013]). *Le foncteur $\mathfrak{M}$ est une équivalence de catégories.*

Leur preuve consiste à montrer que si C est sphériquement complet avec comme groupe de valuations $\mathbb{R}$, alors pour tout groupe rigide analytique de type p-divisible G « hyperbolique » on a $G \simeq \mathring{\mathbb{B}}_C^d$. Ils montrent ensuite que l'on peut descendre du cas C sphériquement complet au cas C quelconque en utilisant le fait que les orbites de Hecke, comme fibres des applications de périodes de de Rham de Rapoport-Zink, sont des espaces rigides de dimension 0.

1.7.3 φ-modules sur A_{inf} et modifications de fibrés. Il s'est alors posé la question de savoir comment faire le lien par un objet intermédiaire entre périodes de Hodge-Tate et de de Rham. Le point de départ est la remarque suivante qui résulte du théorème de factorisation 1.3 : les modifications de $\mathcal{O} \hookrightarrow \mathcal{O}(d)$ sont toutes obtenues par un procédé itératif qui est le suivant. Partant de ξ primitif de degré d tel que $\xi = \sum_{n\geq 0}[x_n]\pi^n$ avec $x_d \in 1 + \mathfrak{m}_F$ on peut former le produit de Weierstrass

$$\prod_{n\geq 0} \frac{\varphi^n(\xi)}{\pi^d} \cdot « \prod_{n<0} \varphi^n(\xi) » \in \mathcal{O}(Y)^{\varphi=\pi^d}$$

le produit entre guillemets n'étant pas convergeant mais pouvant être défini comme solution de l'équation fonctionelle $\varphi(x) = \xi x$ dans $\mathbf{A}$ (cf. Fargues et Fontaine [2017] sec. 6.3). Cela a conduit (Fargues [2015c]) à la « version perfectoïde » suivante des φ-modules de Breuil-Kisin (Kisin [2006]). Soit ξ primitif de degré 1 qui engendre le noyau de $\theta : \mathbf{A} \to \mathcal{O}_C$. Par définition un φ-module sur $\mathbf{A}$ est un couple (M, φ) où M est un $\mathbf{A}$-module libre et $\varphi : M \to M$ est semi-linéaire de conoyau annulé par une puissance de ξ. On a alors le résultat suivant.

Théorème 1.12 (Fargues [2015c], Scholze [2014]). *Il y a une anti-équivalence entre $\varphi-\mathrm{Mod}_{\mathbf{A}}$ et la catégorie des modifications de fibrés en ∞, $\mathcal{F} \hookrightarrow \mathcal{G}$ avec $\mathcal{F}$ trivial muni d'un réseau dans $H^0(X, \mathcal{F})$.*

Partant d'un φ-module (M, φ) on construit un *« Shtuka »* $(\mathfrak{M}, \varphi) = (M \otimes \mathcal{O}_Y, \varphi \otimes \varphi)$ sur Y dont le zéro est localisé en $V(\xi)$, coker$(\mathfrak{M} \xrightarrow{\varphi} \mathfrak{M})$ est supporté en $V(\xi)$. On forme alors (Fargues [2015c] sec. 4.4)

$$\mathfrak{M}_\infty = \varprojlim_{n\geq 0} \varphi^{n*}\mathfrak{M} \ \text{ et } \ \mathfrak{M}^\infty = \varinjlim_{n\leq 0} \varphi^{n*}\mathfrak{M},$$

(opération qui est l'analogue du produit de Weierstrass précédent pour GL_1). Puisque φ « dilate sur le disque épointé Y » les deux limites précédentes sont essentiellement

constantes sur tout ouvert quasicompact de Y, et $\mathfrak{M}_\infty$ et $\mathfrak{M}^\infty$ sont donc des fibrés φ-équivariants i.e. des fibrés sur X^{ad}. La modification associée à (M,φ) est alors la duale de $\mathfrak{M}_\infty \hookrightarrow \mathfrak{M}^\infty$.
Dans l'autre sens une telle modification fournit par un procédé inverse un tel Shtuka $(\mathfrak{M},\varphi)$ sur Y. On utilise alors la compactification par deux diviseurs (Scholze [2014])

$$\mathcal{Y} = \mathrm{Spa}(\mathbf{A},\mathbf{A}) \setminus V(\pi,[\varpi]) = Y \cup V(\pi) \cup V([\varpi]).$$

En utilisant le théorème 11.1.7 de Fargues et Fontaine [2017] on vérifie que $\mathfrak{M}$ s'étend automatiquement grâce à la structure de Frobenius le long du diviseur $([\varpi])$. En utilisant la condition de trivialité de $\mathscr{F}$ couplée au choix du réseau dans $H^0(X,\mathscr{F})$ on vérifie que $\mathfrak{M}$ s'étend en $\pi = 0$. Le fibré $\mathfrak{M}$ s'étend donc canoniquement à $\mathcal{Y}$. On peut alors appliquer le résultat GAGA de Kedlaya (Kedlaya [2016b]) qui dit que les fibrés sur $\mathrm{Spec}(\mathbf{A}) \setminus V(\pi,[\varpi])$ et $\mathrm{Spa}(\mathbf{A},\mathbf{A}) \setminus V(\pi,[\varpi])$ s'identifient et que ceux-ci sont triviaux pour conclure ($\mathbf{A}$ se comporte en quelque sorte comme un anneau local régulier de dimension 2).

Partant de $(M,\varphi) \in \varphi\text{-}\mathrm{Mod}_{\mathbf{A}}$, si $(\mathfrak{N},\varphi) = (M,\varphi)\otimes \mathcal{O}_{\mathfrak{Y}}$:
- si $V(\pi) = \{y_{\text{ét}}\}$ (diviseur étale) : $(\mathfrak{N}_{y_{\text{ét}}},\varphi)$ fournit le fibré trivial $\mathscr{F}^\vee$ ainsi que le réseau dans $H^0(X,\mathscr{F}^\vee)$,
- si $V([\varpi]) = \{y_{\text{cris}}\}$ (diviseur cristallin) : $(\mathfrak{N}_{y_{\text{cris}}},\varphi)$ fournit le fibré $\mathscr{G}^\vee$ qui est associé à l'isocristal $(\mathfrak{N}_{y_{\text{cris}}}\otimes k(y_{cris}),\varphi)$,
- si $V(\xi) = \{y_{\text{dR}}\}$ (diviseur de de Rham) : l'inclusion de B_{dR}^+-réseaux $\varphi : \widehat{\mathfrak{N}}_{\varphi(y_{\text{dR}})} \hookrightarrow \widehat{\mathfrak{N}}_{y_{\text{dR}}}$ fournit la filtration de Hodge i.e. la modification $\mathscr{G}^\vee \hookrightarrow \mathscr{F}^\vee$.

C'est en ce sens là que l'on a unifié les différentes périodes. On renvoie à Scholze [2014] pour plus de détails.

Comme corollaire du théorème précédent on déduit comme dans Kisin [2006] que la catégorie des groupes p-divisibles sur $\mathcal{O}_C$ s'identifie à celle des φ-modules sur $\mathbf{A}$, (M,φ), tels que coker φ soit annulé par ξ (Fargues [2015c] sec 4.8). Ce résultat a été repris par Lau sur des bases affinoïdes perfectoïdes plus générales dans Lau [2016] par les méthodes de Displays/Windows de Zink. Remarquons enfin que sur $\mathcal{O}_C$ l'hypothèse $p \neq 2$ de Fargues [2015c] et Lau [2016] n'est en fait pas nécessaire puisque tout groupe p-divisible $\mathcal{G}$ sur $\mathcal{O}_C$ est somme directe d'un groupe étale et d'un groupe connexe.

Si X est un schéma propre et lisse sur $\mathcal{O}_K$ de valuation discrète à corps résiduel parfait avec $C = \widehat{\overline{K}}$, les différents théorèmes de comparaison cristallins/semi-stables (Fontaine et Messing [1987], Tsuji [1999], Niziol [2008] par exemple) fournissent automatiquement de telles modifications de fibrés avec $\mathscr{F} = H^\bullet_{\text{ét}}(X_{\overline{K}},\mathbb{Q}_p)\otimes \mathcal{O}_X$ et $\mathscr{G} = \mathscr{E}(D,\varphi)$,

(D, φ) étant la cohomologie cristalline de la fibre spéciale X_{k_K}. On peut donc définir grâce à 1.12 une théorie cohomologique sur de tels X à valeurs dans les φ-modules sur $\mathbf{A}$.

À la suite de cela Bhatt, Morrow et Scholze (Bhatt, Morrow et Scholze [2015], Bhatt, Morrow et Scholze [2016]) ont défini de manière géométrique une telle théorie cohomologique sur les schémas formels propres et lisses sur $\mathcal{O}_C$, à valeurs dans des complexes de φ-modules sur $\mathbf{A}$ et ont donné des applications à l'étude de la torsion dans la cohomologie étale de ces variétés. Le point est qu'ils ne construisent pas seulement des φ-modules sur $\mathbf{A}$ mais plutôt un *complexe de cohomologie*, ce qui raffine la construction précédente et donne des informations sur la torsion. En effet, l'équivalence du théorème 1.12 n'est pas une équivalence exacte et ne passe pas aux complexes de cohomologie.

Notons également que Niziol a interprété la cohomologie syntomique géométrique en termes de fibrés sur la courbe dans Niziol [p. d.]. Plus précisément, la cohomologie de Deligne s'interprète parfois comme groupes d'extensions de structures de Hodge mixtes, les régulateurs archimédiens envoyant une extension de motifs mixtes sur l'extension correspondante de structures de Hodge. Niziol donne une interprétation similaire dans Niziol [ibid.] pour la cohomologie syntomique en termes d'extensions de modifications de fibrés vectoriels.

Enfin concluons par la conjecture suivante, apparue au cours de nombreuses discussions avec Le Bras et Scholze, également inspirée par les travaux de Colmez et Niziol (Colmez et Niziol [2017] sec. 5.2 où les espaces de Banach-Colmez apparaissent naturellement) qui devrait généraliser la cohomologie de Hyodo-Kato des schémas propres et lisses à réduction semi-stable.

Conjecture 1.13. *Soit $C|\mathbb{Q}_p$ complet et algébriquement clos. On peut définir une théorie cohomologique sur les C-espaces rigides lisses quasicompacts séparés, à valeurs dans les fibrés vectoriels sur la courbe associée à $C^\flat$ qui généralise la cohomologie de Hyodo-Kato des schémas propres et lisses à réduction semi-stable. Cette théorie cohomologique devrait se factoriser par la catégorie des motifs rigides de Ayoub (Ayoub [2015]).*

On espère même la construction d'un complexe de cohomologie dans $\mathbb{D}^b_{coh}(\mathcal{O}_X)$. Par semi-simplification de la filtration de Harder-Narasimhan cela fournirait une théorie cohomologique à valeurs dans les isocristaux. Cependant la conjecture précédente est plus subtile puisqu'on prédit l'existence d'un relèvement au niveau des fibrés. Bien sûr si $C = \widehat{\overline{K}}$ cela devrait définir une théorie cohomologique sur de tels K-espaces rigides à valeurs dans les fibrés Galois équivariants.

2 Géométrisation de l'ensemble de Kottwitz et applications

2.1 Géométrisation. Dans cette section X est la courbe associée à un corps F algébriquement clos et E comme précédemment. Soit G un groupe réductif sur E. On considère l'ensemble de Kottwitz $B(G) = G(\breve{E})/\sigma$-conj. des classes d'isomorphismes de G-isocristaux. Ici par σ-conjugaison on entend $b \sim gbg^{-\sigma}$. Si $b \in G(\breve{E})$ on peut lui associer un G-fibré sur X par composition

$$\begin{aligned} \mathrm{Rep}(G) &\longrightarrow \varphi\mathrm{-Mod}_{\breve{E}} \xrightarrow{\mathcal{E}(-)} \text{fibrés sur } X \\ (V,\rho) &\longmapsto \big(V_{\breve{E}}, \rho(b)\sigma\big). \end{aligned}$$

On le note $\mathcal{E}_b$ et on le voit également comme un G-torseur étale. Tiré en arrière sur la courbe adique il s'agit du G^{an}-torseur localement trivial pour la topologie adique

$$Y \times_{\varphi^{\mathbb{Z}}} G^{an}_{\breve{E}}$$

où φ agit sur $G_{\breve{E}}$ via $b\sigma$. Le théorème principal de Fargues [2016a] est alors le suivant (cf. Anschütz [2017] pour le cas $E = \mathbb{F}_q((\pi))$). Il s'agit d'une généralisation du théorème de classification 1.5.

Théorème 2.1 (Fargues [2016a], Anschütz [2017]). *Il y a une bijection d'ensembles pointés*

$$\begin{aligned} B(G) &\xrightarrow{\sim} H^1_{\text{ét}}(X,G) \\ [b] &\longmapsto [\mathcal{E}_b]. \end{aligned}$$

Ce résultat est utilisé dans Caraiani et Scholze [2017] afin de construire des applications de périodes de Hodge-Tate pour des variétés de Shimura de type Hodge et stratifier par un ensemble de Kottwitz la variété de drapeaux au but de ces applications de périodes.

Cette bijection satisfait des propriétés très agréables :

- Il y a un *dictionnaire entre théorie de la réduction de Harder-Narasimhan/ Atiyah-Bott et la description par Kottwitz de* $B(G)$. Par exemple :
  - b est basique ssi $\mathcal{E}_b$ est semi-stable
  - le polygone de Harder-Narasimhan $\{\nu_{\mathcal{E}_b}\}$ vu comme classe de conjugaison géométrique de cocaractère $\mathbb{D} \to G_{\overline{F}}$ est égal à $\{\nu_b^{-1}\}$ l'inverse du polygone de Newton.
- L'application $\kappa : B(G) \to \pi_1(G)_\Gamma$ de Kottwitz s'interprète comme l'opposé d'une *première classe de Chern d'un G-fibré.*
- On dispose d'un *analogue du théorème de Drinfeld-Simpson* : lorsque G est quasi-déployé, pour tout point fermé $\infty \in |X|$, $\mathcal{E}_{|X\setminus\{\infty\}}$ est trivial. Les fibrés sur X s'obtiennent ainsi par recollement « à la Beauville-Laszlo » et donc si $\{\infty\} = V^+(t)$, de corps résiduel C, le groupoïde des G-fibrés sur X s'identifie au groupoïde quotient

$$\big[G\big(\mathcal{O}(Y)[\tfrac{1}{t}]^{\varphi=Id}\big)\backslash G(B_{dR}(C))/G(B^+_{dR}(C))\big].$$

2.2 Interprétation géométrique de certains résultats de Tate. On suppose dans cette section que $E|\mathbb{Q}_p$. Durant la démonstration du théorème 2.1 est apparu un lien intéressant entre théorie du corps de classe et fibrés sur la courbe. Par exemple, si B est une algèbre à division sur E, d'après la théorie du corps de classe (calcul du groupe de Brauer), il existe un isocristal isocline (D, φ) tel que $B \simeq End(D, \varphi)$. On a alors

$$B \otimes_E \mathcal{O}_X \xrightarrow{\sim} \mathcal{E}nd(\mathcal{E}(D, \varphi)).$$

Il en résulte que le morphisme $\mathrm{Br}(E) \to \mathrm{Br}(X)$ est nul. Partant de ce résultat on démontre dans Fargues [2016a] que *le groupe de Brauer* $\mathrm{Br}(X)$ *est nul.* On conjecture même en fait que le corps des fonctions de X est (C1) (Fargues [ibid.]).

On peut aller plus loin dans les liens entre corps de classe et cohomologie de la courbe. On note $\Gamma = \mathrm{Gal}(\overline{E}|E)$. D'après le théorème 1.8 les systèmes locaux étales finis sur X correspondent par tiré en arrière via $X \to \mathrm{Spec}(E)$ aux Γ-modules discrets finis. On a alors le résultat suivant.

Théorème 2.2 (Fargues [ibid.]). *Soit M un Γ-module discret fini et $\mathcal{F}$ le faisceau étale localement constant associé sur X par tiré en arrière via $X \to \mathrm{Spec}(E)$. On a alors :*

1. $H^\bullet(E, M) \xrightarrow{\sim} H^\bullet_{\text{ét}}(X, \mathcal{F})$.
2. *Via cet isomorphisme pour $M = \mathbb{Z}/n\mathbb{Z}(1)$, la classe fondamentale de la théorie du corps de classe correspond à la classe fondamentale de la courbe $\eta_X = c_1(\mathcal{O}(1)) \in H^2_{\text{ét}}(X, \mathbb{Z}/n\mathbb{Z}(1))$, c'est à dire la classe de cycle d'un point fermé sur la courbe.*

On peut ainsi réinterpréter (mais pas redémontrer complètement jusqu'à maintenant i.e. par des méthodes purement géométriques) géométriquement les deux théorèmes suivants de Tate :

1. La dualité de Tate-Nakayama s'interprète comme la *dualité de Poincaré sur la courbe.*
2. La formule de Tate pour la caractéristique d'Euler-Poincaré de la cohomologie galoisienne s'interprète comme une *formule de Grothendieck-Ogg-Shafarevich.*

On espère pouvoir étudier les systèmes locaux sur des ouverts de la courbe ainsi que leur cohomologie (cf. Fargues [ibid.] sec. 3.3 pour des conjectures précises et également Fargues [2015a] sec. 7 pour une description conjecturale du groupe fondamental d'un ouvert).

2.3 Application aux espaces de périodes p-adiques. Rapoport a donné une première application du théorème 2.1 aux espaces de périodes de Rapoport et Zink [p. d.] dans

RAPOPORT [p. d.] en montrant que le lieu admissible coïncide avec toute la variété de drapeaux uniquement dans le cas Lubin-Tate et son dual de Cartier. Il faut faire attention ici que, bien que le théorème de Colmez-Fontaine dise que « admissible est équivalent à faiblement admissible », du point de vue géométrique des espaces de périodes cela dit uniquement que les lieux admissible et faiblement admissible ont mêmes points à valeurs dans des corps p-adiques de valuation discrète, mais ils ne sont pas égaux en général.

Rappelons le contexte de ce type de problème (RAPOPORT et ZINK [p. d.],DAT, ORLIK et RAPOPORT [2010]). On fixe une classe de conjugaison géométrique de cocaractère minuscule $\{\mu\}$ à valeurs dans G et on regarde la variété de drapeaux associée $\mathfrak{F}(G,\mu)$ comme espace adique sur le complété de l'extension maximale non ramifiée du corps de définition de $\{\mu\}$. On fixe un isocristal avec G-structure $[b] \in B(G,\mu)$, l'ensemble de Kottwitz (KOTTWITZ [1997]) qui paramètre les strates de Newton dans les variétés de Shimura dont la donnée induit (G,μ) localement en p (la non vacuité des strates de Newton associées à un élément de $B(G,\mu)$, conjecturée dans FARGUES [2004], est connue dans de très nombreux cas maintenant). On regarde le lieu faiblement admissible $\mathfrak{F}(G,\mu,b)^{\mathrm{fa}}$, un ouvert partiellement propre obtenu en enlevant un nombre fini de $J_b(E)$-orbites de variétés de Schubert où la condition de faible admissibilité de Fontaine n'est pas satisfaite. Rapoport et Zink avaient conjecturé dans RAPOPORT et ZINK [p. d.] l'existence d'un ouvert

$$\mathfrak{F}(G,\mu,b)^{\mathrm{a}} \subset \mathfrak{F}(G,\mu,b)^{\mathrm{fa}}$$

ayant même points classiques de Tate (à valeurs dans un corps de valuation discrète) que $\mathfrak{F}(G,\mu)^{\mathrm{a}}$, ainsi que l'existence d'un E-système local étale avec G-structure dessus qui interpolerait les représentations cristallines à valeurs dans $G(E)$ en les points classiques. Ces dernières représentations sont fournies par Fontaine via le théorème « faiblement admissible équivalent à admissible » en ces points. L'existence de cet ouvert est maintenant connue grâce à la courbe, les travaux de Kedlaya-Liu (KEDLAYA et LIU [2015], cf. le point (1) du théorème 3.1) et de Scholze (SCHOLZE [p. d.]), cf. sec. 3 de CHEN, FARGUES et SHEN [2017] pour un rapide survol. Grâce à cela on peut construire les variétés de Shimura locales associées (RAPOPORT et VIEHMANN [2014]) comme espace de réseaux dans ce système local (de JONG [1995]).

Dans HARTL [2013] Hartl classifie les μ minuscules associés à $G = \mathrm{GL}_n$ tels que $\mathfrak{F}^{\mathrm{a}} = \mathfrak{F}^{\mathrm{fa}}$. Inspirés entre autre par ce résultat, Rapoport et l'auteur ont conjecturé le résultat suivant qui est maintenant démontré.

Théorème 2.3 (CHEN, FARGUES et SHEN [2017])**.** *Pour $[b] \in B(G,\mu)$ basique sont équivalents*

1. $\mathfrak{F}(G,\mu,b)^{\mathrm{a}} = \mathfrak{F}(G,\mu,b)^{\mathrm{fa}}$.
2. *L'ensemble $B(G,\mu)$ est pleinement HN décomposable.*

Par définition l'ensemble $B(G,\mu)$ est pleinement HN décomposable si pour tout $[b'] \in B(G,\mu)$ non basique son polygone de Newton, comme élément d'une chambre de Weyl positive, « touche le polygone de Hodge défini par μ en dehors de ses extrémités ». Les espaces/variétés de Shimura associés aux ensembles $B(G,\mu)$ pleinement HN décomposables jouissent de propriétés tout à fait remarquables (Goertz, He et Nie [2016]). Le premier exemple de tel ensemble qui a intrigué l'auteur remonte à « l'astuce de Boyer » aka la décomposition de Hodge-Newton (Boyer [1999] pour l'astuce originelle, Harris et Taylor [2001] pour sa variante variétés de Shimura, Mantovan [2008], Mantovan et Viehmann [2010], Shen [2014b] pour des généralisations au cas des variétés de Shimura, Goertz, He et Nie [2016] pour le cas de la fibre spéciale, Hansen [2016] and Gaisin et Imai [2016] pour des versions "modernes" dans le cadre des espaces de module de Shtukas locaux).

Voici un exemple de $(2) \Rightarrow (1)$ dans le théorème précédent. Il avait été traité auparavant (Fargues [2015b], appendice de Shen [2016]) et a servi de guide dans la démonstration de $(2) \Rightarrow (1)$.

Corollaire 2.4 (Application de la courbe aux surfaces)**.** *Supposons $p \neq 2$. Soit $\mathcal{F} = \{q = 0\} \subset \P^{20}_{\breve{\mathbb{Q}}_p}$ la variété de drapeaux formées des droites isotropes pour la forme quadratique q telle que $q(x) = \sum_{i=1}^{21} x_i x_{22-i}$. Soit $Z = \{x_{12} = \cdots = x_{21} = 0\} \cap \mathcal{F}$, une variété de Schubert pour $G = SO(q)$. L'espace des périodes p-adiques des surfaces K3 polarisées à réduction supersingulière s'identifie à $\mathcal{F} \setminus G(\mathbb{Q}_p) \cdot Z$.*

La démonstration du théorème 2.3 repose sur la construction suivante. À $x \in \mathcal{F}(G,\mu)(C)$ est associé une modification $\mathcal{E}_{b,x}$ de $\mathcal{E}_b$. En effet, $\mathcal{E}_b$ est canoniquement trivialisé en ∞ et on utilise le fait que μ est minuscule afin d'associer à un tel x une donnée de modification dans $G(B_{dR})/G(B_{dR}^+)$. Via le théorème 2.1 la classe d'isomorphisme de $\mathcal{E}_{b,x}$ fournit un élément de $B(G)$. On peut classifier ces éléments possibles par un autre ensemble (fini) de Kottwitz $B(G, 0, \nu_b\mu^{-1})$ (Rapoport [p. d.] coro. 0.10, Chen, Fargues et Shen [2017] sec. 4). Cela fournit une stratification de $\mathcal{F}(G,\mu)$, la strate ouverte étant celle associée à $[1] \in B(G)$ qui est exactement l'ouvert admissible i.e. le lieu où le G-fibré $\mathcal{E}_{b,x}$ est trivial (cf. Chen, Fargues et Shen [ibid.] sec. 5). C'est l'étude de cette stratification qui mène à la preuve du théorème 2.3. Ce type de stratification apparaît également dans Caraiani et Scholze [2017] par le même type de construction. En effet, les périodes de Hodge-Tate et de de Rham de la Section 1.7.2 s'interprètent agréablement au sens où si $\mathcal{E}_b \rightsquigarrow \mathcal{E}_1$ est une modification minuscule de type μ en ∞ :

— sa *période de de Rham* est donnée par l'élément $x \in \mathcal{F}(G,\mu)(C)$ qui identifie la modification à $\mathcal{E}_b \rightsquigarrow \mathcal{E}_{b,x}$

— sa *période de Hodge-Tate* est donnée par l'élément $y \in \mathcal{F}(G,\mu^{-1})(C)$ qui identifie la modification à $\mathcal{E}_{1,y} \rightsquigarrow \mathcal{E}_b$.

Ces périodes de Hodge-Tate et de de Rham sont l'ombre de deux pattes du champ de Hecke des modifications des G-fibrés (cf. Section 3.3). Ce type de considérations sur les périodes de Hodge-Tate et de de Rham a été une très forte inspiration pour l'introduction de la conjecture de géométrisation de la Section 3.

Puisque nous l'utilisons dans la preuve du théorème 2.3, remarquons que la considération des G-fibrés sur la courbe, pour G quelconque, clarifie complètement *l'isomorphisme entre les tours jumelles* de Faltings [2002] et Faltings [2004] (cf. également Fargues [2008] et Scholze et Weinstein [2013]). Plus précisément, pour $[b] \in B(G)$ basique le groupe réductif J_b devient une forme intérieure pure de G après extension des scalaires à la courbe : *$J_b \times X$ est la torsion intérieure de $G \times X$ par le G-torseur $\mathcal{E}_b$.* Il en résulte une équivalence de groupoïdes entre J_b-fibrés sur X et G-fibrés sur X qui respecte les modifications. À partir de là les espaces de modules de modifications sont identifiés. On renvoie à la section 5.1 de Chen, Fargues et Shen [2017] pour plus de détails.

Enfin, l'auteur a commencé à s'intéresser aux filtrations du type Harder-Narasimhan en théorie de Hodge p-adique lors de la découverte de l'existence de telles filtrations sur les schémas en groupes finis et plats (Fargues [2010b]). Ces filtrations permettent de définir des domaines fondamentaux dans les espaces de modules de groupes p-divisibles (Fargues [ibid.] coro. 11 et Shen [2014a]). On a alors la conjecture suivante concernant l'existence de « domaines fondamentaux de Siegel » dans les espaces de périodes p-adiques. Cette conjecture est liée au théorème 2.3 via le fait que lorsque $B(G,\mu)$ est pleinement HN décomposé alors $\mathcal{F}(G,\mu) \setminus \mathcal{F}(G,\mu,b)^{\mathrm{a}}$ est « paraboliquement induit » (Chen, Fargues et Shen [2017] sec. 7).

Conjecture 2.5 (Chen, Fargues et Shen [ibid.] sec. 7)**.** *Pour $[b] \in B(G,\mu)$ basique, sont équivalents :*

1. $\mathcal{F}(G,\mu,b)^{\mathrm{a}} = \mathcal{F}(G,\mu,b)^{\mathrm{fa}}$
2. *Il existe un ouvert quasicompact $U \subset \mathcal{F}(G,\mu,b)^{\mathrm{a}}$ tel que $\mathcal{F}(G,\mu,b)^{\mathrm{a}} = J_b(E)\cdot U$.*

Si $\underline{G}$ est un modèle entier de G, les filtrations de Harder-Narasimhan des schémas en groupes finis et plats s'étendent aux φ-modules de Breuil-Kisin avec $\underline{G}$-structure (Levin et Wang-Ericksson [2016], Peche Irissarry [p. d.]) ainsi qu'aux φ-modules sur A_{inf} de la Section 1.7.3. En étudiant les $\underline{G}$-Shtukas sur $\mathfrak{Y}$ il est probable que l'on puisse généraliser la preuve de 1.12 à cadre là. On peut donc espérer des constructions générales de tels domaines fondamentaux grâce à ces techniques.

3 Géométrisation de la correspondance de Langlands locale

3.1 La courbe en familles. Soit $\mathrm{Perf}_{\mathbb{F}_q}$ la catégorie des $\mathbb{F}_q$-espaces perfectoïdes (Scholze [2012]). On ne s'intéresse désormais qu'à la courbe adique et on note X ce que l'on notait auparavant X^{ad}. Pour $S \in \mathrm{Perf}_{\mathbb{F}_q}$ on peut construire un E-espace adique X_S que l'on peut voir comme étant « la famille de courbes adiques $(X_{k(s),k(s)^+})_{s\in S}$ ». Comme dans le cas d'un point base

$$X_S = Y_S/\varphi^{\mathbb{Z}}$$

où Y_S est « Stein ». Par exemple, si $E = \mathbb{F}_q((\pi))$, $Y_S = \mathbb{D}_S^* = \{0 < |\pi| < 1\} \subset \mathbb{A}_S^1$ et le Frobenius φ est donné par celui de S. Lorsque $S = \mathrm{Spa}(R, R^+)$ est affinoïde perfectoïde l'espace Y_S se définit comme étant

$$\mathrm{Spa}(\mathbf{A}, \mathbf{A}) \setminus V(\pi[\varpi])$$

où $\mathbf{A} = W_{\mathcal{O}_E}(R^+)$ si $E|\mathbb{Q}_p$, $\mathbf{A} = R^+[\![\pi]\!]$ si $E = \mathbb{F}_q((\pi))$ et ϖ est une pseudo-uniformisante de R, $0 < |\varpi| < 1$. Ces E-espaces adiques sont pré-perfectoïdes et on montre que

$$Y_S^{\diamond} = S \times \mathrm{Spa}(E)^{\diamond}$$

où $\varphi \leftrightarrow \varphi_S \times Id$. Ici on utilise la notion de diamant introduite par Scholze (Scholze [2014], Scholze [p. d.]). Si Z est un E-espace adique alors $Z^{\diamond}$ est le faisceau sur $\mathrm{Perf}_{\mathbb{F}_q}$ tel que

$$Z^{\diamond}(T) = \{(T^{\sharp}, \iota, f) \,|T^{\sharp} \text{ est perfectoïde}, \iota : T \xrightarrow{\sim} T^{\sharp,\flat}, f : T^{\sharp} \to Z\}/\sim .$$

Cette formule catégorique, bien que particulièrement élégante, ne dit rien sur la géométrie de Y_S (par exemple elle ne dit rien sur ce qu'est un fibré vectoriel sur X_S).

3.2 Résultats de Kedlaya et Liu sur les familles de fibrés. Kedlaya et Liu ont montré que l'on dispose d'une bonne notion de fibré vectoriel sur des espaces adiques du type X_S (Kedlaya et Liu [2015] sec. 2.7). Ils ont également démontré les trois résultats suivants qui sont au coeur de la structure du champ des G-fibrés sur la courbe (leurs résultats sont énoncés de manière moins « géométrique » en termes de φ-modules sur des anneaux de Robba mais sont équivalents à ceux qui suivent). Dans cet énoncé

$$\tau : \widetilde{(X_S)}_{\text{pro-ét}} \longrightarrow \widetilde{S}_{\text{pro-ét}}$$

est un morphisme de topos pro-étales obtenu par exemple grâce à la fonctorialité de la courbe en la base S.

Théorème 3.1. *Pour S un $\mathbb{F}_q$-espace perfectoïde et $\mathcal{E}$ un fibré vectoriel sur X_S*

1. *La fonction $|S| \ni s \longmapsto \mathrm{HN}(\mathcal{E}_{|X_{k(s),k(s)^+}})$ (polygone de Harder-Narasimhan) est semi-continue supérieurement. En particulier le lieu semi-stable dans S est un ouvert partiellement propre.*

2. *Les foncteurs $R\tau_*$ et $\tau^*(-)\otimes_{\underline{E}}\mathcal{O}_X$ induisent des équivalences inverses entre fibrés vectoriels, fibre à fibre sur S semi-stables de pente 0, et E-systèmes locaux pro-étales sur S.*
3. *Le fibré $\mathcal{O}(1)$ est ample au sens où, si S est affinoïde perfectoïde, alors pour $d \gg 0$ il existe une surjection $\mathcal{O}_{X_S}^n \twoheadrightarrow \mathcal{E}(d)$.*

Le point (1) se reformule de façon agréable de la façon suivante (c'est un exercice de vérifier que les deux sont équivalents). Il y a deux fonctions additives (introduites par Colmez dans Colmez [2002]) sur la catégorie des espaces de Banach-Colmez sur $C|E$ (cf. Section 1.6). Ce sont les fonctions dimension et hauteur déterminées par $\dim \mathbb{G}_a = 1$, $\dim \underline{E} = 0$, $\mathrm{ht}\, \mathbb{G}_a = 0$ et $\mathrm{ht} \underline{E} = 1$ (par référence à la hauteur et la dimension du revêtement universel d'un groupe p-divisible, un cas particulier d'espace de Banach-Colmez). Pour un fibré vectoriel $\mathcal{F}$ sur la courbe X_F associée à $F = C^\flat$ on peut alors définir la dimension et la hauteur des espaces de Banach-Colmez $H^0(X_F, \mathcal{F})$ et $H^1(X_F, \mathcal{F})$ (plus précisément, ces espaces sont définis par troncature via la t-structure du théorème 1.9, la fonction dimension sur le coeur de cette t-structure coïncide alors avec le degré et la hauteur avec le rang, les deux vues comme fonctions additives sur $\mathbb{D}^b_{coh}(\mathcal{O}_{X_F})$). Le point (1) se reformule alors en le fait que pour $\bullet \in \mathbb{N}$, la fonction

$$|S| \longrightarrow \big(\dim H^\bullet(X_{k(s),k(s)^+}, \mathcal{E}_{|X_{k(s),k(s)^+}}), \mathrm{ht}\, H^\bullet(X_{k(s),k(s)^+}, \mathcal{E}_{|X_{k(s),k(s)^+}})\big) \in \mathbb{N}\times\mathbb{Z}$$

est semi-continue supérieurement (puisque $\dim H^0 - \dim H^1 = \deg \mathcal{E}$ et $\mathrm{ht}\, H^0 - \mathrm{ht}\, H^1 = \mathrm{rg} \mathcal{E}$ qui sont localement constants, cela se ramène au même énoncé pour le H^0). Cela replace ce type de résultat dans le cadre conceptuel des résultats de semi-continuité de Grothendieck [1963] (mais malheureusement l'auteur ne sait pas donner un sens au fait que $R\tau_*\mathcal{E}$ soit un « complexe parfait de quoi que ce soit », ce qui donnerait une preuve rapide de cette semi-continuité).

Le point (2) est une vaste généralisation aux $\mathbb{Q}_p$-systèmes locaux du théorème de Lang (qui concerne les $\mathbb{Z}_p$-systèmes locaux). Plus précisément, si $S = \mathrm{Spa}(R, R^+)$ alors d'après Lang,

$$\underline{\mathcal{O}_E}\text{-systèmes locaux pro-étales} \xrightarrow{\sim} \varphi\text{−}\mathrm{Mod}^{\text{ét}}_{\mathbf{A}_{R,R^+}}$$

où les φ-modules sont ceux de la Section 1.7.3 et la condition d'être étale signifie simplement que $\varphi : M \xrightarrow{\sim} M$ est un isomorphisme. Cette condition se réécrit du point de vue du théorème précédent en demandant que pour tout $s \in S$, $(M, \varphi) \otimes_{\mathbf{A}_{R,R^+}} \mathbf{A}_{k(s),k(s)^+}$ soit étale. Le point (2) du théorème précédent remplace donc la condition d'être étale fibre à fibre par la condition d'être semi-stable de pente 0 fibre à fibre et l'anneau $\mathbf{A}$ par un anneau de Robba.

Remarquons que le point (2) est également une généralisation du résultat « du type Narasimhan-Seshadri » 1.6.

Pour le point (3) on renvoie à la discussion après 1.7.

3.3 Le champ Bun_G **et la conjecture de géométrisation.** Motivé par l'apparition systématique des modifications de fibrés vectoriels (Section 1.7.2), le théorème 2.1, les conjectures de Kottwitz décrivant la partie discrète de la cohomologie des espaces de Rapoport-Zink (Rapoport [1995]), les travaux naissants de Scholze sur la construction de L-paramètres via la cohomologie des espaces de Shtukas locaux (Scholze [2014],Lafforgue [p. d.]) et bien sûr ceux de l'école russe autour du programme de Langlands géométrique (Drinfeld [1983], Gaitsgory [2002]), on a formulé une conjecture de géométrisation de la correspondance de Langlands locale (Fargues [2016b], Fargues et Scholze [p. d.]). Puisqu'il s'agit d'un travail en cours nous n'allons pas nous étendre en détail dessus. Indiquons seulement quelques points en lien avec les résultats précédents.

On définit le champ Bun_G sur $\mathrm{Perf}_{\overline{\mathbb{F}}_q}$ comme étant $S \mapsto \{G\text{-fibrés sur } X_S\}$. On le voit comme un champ pour la v-topologie (Scholze [p. d.]). Le point de départ est que le théorème 2.1 donne une identification

$$B(G) = |\mathrm{Bun}_G|.$$

Il faut prendre garde au fait que la topologie quotient sur $B(G) = G(\breve{E})/\sigma$-conj. est la topologie discrète (si les matrices de Frobenius de deux isocristaux sont proches elles sont σ-conjuguées). *Ce n'est pas la topologie qui nous intéresse, on regarde plutôt celle de* $|\mathrm{Bun}_G|$ *induite par les sous-champs ouverts.* L'application première classe de Chern , i.e. $-\kappa$ sur $B(G)$ (cf. Section 2.1), est localement constante (Fargues et Scholze [p. d.]) et fournit une décomposition en sous-champs ouverts/fermés

$$\mathrm{Bun}_G = \coprod_{\alpha \in \pi_1(G)_\Gamma} \mathrm{Bun}_G^\alpha.$$

Il y a une application polygone de Harder-Narasimhan semi-continue à valeurs dans une chambre de Weyl (cet énoncé peut se déduire de 3.1 (1)). Contentons nous seulement de dire que *le lieu semi-stable est ouvert*. Cet énoncé est à mettre en parallèle avec le fait que le lieu basique dans la fibre spéciale d'une variété de Shimura est fermé ; la cohérence entre les deux énoncés provenant du fait que le tube, au sens de la géométrie rigide, au dessus d'un fermé est ouvert. D'après Kottwitz (Kottwitz [1985]) la restriction de κ induit une bijection $B(G)_{basique} \xrightarrow{\sim} \pi_1(G)_\Gamma$. Cela s'interprète géométriquement en : *toute composante de* Bun_G *indexée par un élément de* $\pi_1(G)_\Gamma$ *possède un unique point semi-stable* (l'auteur n'avait jamais vraiment compris la signification de cet énoncé de Kottwitz avant de tomber sur ce simple énoncé géométrique). Soit donc $[b]$ basique et $\alpha = -\kappa(b)$. En utilisant le point (2) de 3.1 on peut démontrer le résultat suivant qui calcule la gerbe résiduelle en l'unique point semi-stable de la composante associée à α :

via le G-fibré $\mathcal{E}_b$, il y a un isomorphisme

$$[\mathrm{Spa}(\overline{\mathbb{F}}_q)/\underline{J_b(E)}] \xrightarrow{\sim} \mathrm{Bun}_G^{\alpha,ss}.$$

(champ classifiant des $J_b(E)$-torseurs pro-étales). *C'est là une des différences majeures avec le champ « classique » des G-fibrés sur une courbe : le groupe des automorphismes du G-fibré trivial est $\underline{G(E)}$ qui est totalement discontinu tandis que dans la situation « classique » c'est le groupe algébrique connexe G.* Cette différence de nature entre champs classifiants, couplée aux travaux de Kaletha et Kottwitz (KALETHA [2016] par exemple), est importante afin de comprendre la conjecture de géométrisation. Plus précisément, si l'on prend la fibre en un point semi-stable d'un faisceau ℓ-adique sur Bun_G on obtient une représentation lisse de $J_b(E)$ (alors que dans la situation « classique », puisque G est connexe, on n'obtiendrait qu'un simple espace vectoriel ℓ-adique).

Venons en maintenant aux modifications de fibrés qui ont joué un rôle très important dans l'intuition de la conjecture. Ici l'intuition vient du point (2) de 1.1. En effet, les débasculements de $S \in \mathrm{Perf}_{\overline{\mathbb{F}}_q}$ sur E, $S^\sharp$ un E-espace perfectoïde avec $S \xrightarrow{\sim} S^{\sharp,\flat}$, fournissent des diviseurs de Cartier

$$S^\sharp \hookrightarrow X_S$$

donnés par $S^\sharp = V(\xi) \hookrightarrow Y_S$ si S est affinoïde perfectoïde et $\xi \in \mathbf{A}_{R,R^+}$ est primitif de degré 1 définissant le débasculement $S^\sharp$. Posons alors

$$\mathrm{Div}^1_S = S \times \mathrm{Spa}(E)^\diamond/\varphi_{E^\diamond}^{\mathbb{Z}}$$

comme S-diamant. Cet objet est en quelque sorte « le miroir » de $X_S^\diamond = S \times \mathrm{Spa}(E)^\diamond/\varphi_S^{\mathbb{Z}}$ (tous deux ont même site étale puisque $\varphi_S \circ \varphi_{E^\diamond}$ est le Frobenius absolu de $S \times \mathrm{Spa}(E)^\diamond$, c'est ce que l'on utilise dans la Section 1.5). Le faisceau Div^1 est alors celui des « diviseurs de Cartier effectifs de degré 1 sur la courbe » (qui n'est pas la courbe elle-même contrairement à la situation « classique »).

Un autre point important est la construction d'un *système de carte perfectoïdes « lisses »* (ℓ-cohomologiquement lisses au sens de SCHOLZE [p. d.] pour être plus précis) sur le champ Bun_G, dont on montre au final qu'il est lisse de dimension 0 (FARGUES et SCHOLZE [p. d.]). Le point de départ a été la remarque suivante liée aux espaces de Banach-Colmez (Section 1.6) : le champ classifiant $[\mathrm{Spa}(\overline{\mathbb{F}}_q)/\underline{\mathrm{GL}_n(\mathbb{Q}_p)}]$ est lisse. En effet, l'espace de module des injections $\mathcal{O}^n \hookrightarrow \mathcal{O}(1)^n$ s'identifie à l'ouvert U du diamant relatif $H^0(\mathcal{O}(1))^{n^2} \to \mathrm{Spa}(\overline{\mathbb{F}}_q)$ formé des matrices de déterminant dans $H^0(\mathcal{O}(n)) \setminus \{0\}$. L'espace de Banach-Colmez relatif $H^0(\mathcal{O}(1)) \to \mathrm{Spa}(\overline{\mathbb{F}}_q)$ est représentable par un disque ouvert perfectoïde (le revêtement universel d'un groupe de Lubin-Tate) et $U \to \mathrm{Spa}(\overline{\mathbb{F}}_q)$ est donc lisse. Le groupe $\underline{\mathrm{GL}_n(\mathbb{Q}_p)}$ agit librement sur U, si $K \subset \mathrm{GL}_n(\mathbb{Q}_p)$ est un sous-groupe compact ouvert pro-p et $F = \overline{\mathbb{F}}_q((T^{1/p^\infty}))$ alors

$$U_F/\underline{K} \longrightarrow \left[\mathrm{Spa}(\overline{\mathbb{F}}_q)/\underline{\mathrm{GL}_n(\mathbb{Q}_p)}\right]$$

est une présentation lisse par un diamant lisse.

Finalement, on peut alors définir des correspondances de Hecke pour μ une classe de conjugaison de cocaractère

$$\begin{array}{ccc} & \mathrm{Hecke}_\mu & \\ \overleftarrow{h} \swarrow & & \searrow \overrightarrow{h} \\ \mathrm{Bun}_G & & \mathrm{Bun}_G \times \mathrm{Div}^1. \end{array}$$

où $\overrightarrow{h}$ est une fibration étale localement triviale en la cellule de Schubert associée à μ dans la B_{dR}-*Grassmanienne de Scholze* au dessus de Div^1 (Scholze [2014], Caraiani et Scholze [2017]). L'étude des modifications de fibrés en lien avec les groupes p-divisibles (Section 1.7.2) et le lien avec les espaces de périodes (Section 2.3) sont une motivation importante pour l'introduction de ce type de diagramme.

La conjecture se formule alors sommairement de la façon suivante. On suppose G quasi-déployé et on fixe une donnée de Whittaker (la construction devrait dépendre de ce choix). On considère le L-groupe ${}^LG = \widehat{G} \rtimes W_E$ de G sur $\overline{\mathbb{Q}}_\ell$ construit canoniquement à partir d'un isomorphisme de Satake géométrique associé à la B_{dR}-grassmanienne affine de Scholze (Fargues et Scholze [p. d.]). Considérons un L-paramètre discret $\varphi : W_E \to {}^LG$. On conjecture alors l'existence d'un faisceau pervers $\mathscr{F}_\varphi$ sur Bun_G muni d'une action de S_φ, propre pour l'action des correspondances de Hecke et dont les fibres en les points semi-stables de Bun_G réalisent des correspondances de Langlands locales pour toutes les formes intérieures pures de G (les J_b lorsque b parcourt les éléments basiques). Sans rentrer dans les détails, outre le fait que cette conjecture construise des correspondances de Langlands locales dans la direction

$$\text{L-paramètre} \longmapsto \text{représentation},$$

l'un de ses points forts est qu'*elle prédit la structure interne des L-paquets associés à φ via l'action de S_φ*. La propriété de Hecke spécialisée aux points semi-stables *implique automatiquement les conjectures de Kottwitz sur la cohomologie des espaces de Rapoport-Zink* (Rapoport [1995]) via le lien entre modifications de fibrés vectoriels et groupes p-divisibles (Section 1.7.2).

On renvoie à FARGUES [2016b] et FARGUES et SCHOLZE [p. d.] pour un énoncé précis de la conjecture et ses conséquences, en espérant avoir expliqué au lecteur quelques éléments de la démarche qui a mené à sa formulation.

3.4 Simple connexité des fibres d'une application d'Abel-Jacobi. Les résultats présentés dans l'article FARGUES [2017] sont un cas particulier de la conjecture de géométrisation qui ne nécessite pas en quelques sortes de l'avoir comprise entièrement. Il s'agit du cas de GL_1 i.e. de la *théorie du corps de classe géométrique.* Dans ce cas là, dans le cadre « classique », il est bien connu que cette même théorie résulte de ce qu'en grand degré le morphisme d'Abel-Jacobi est une fibration localement triviale en variétés algébriques simplement connexes (des espaces projectifs). On démontre que c'est également le cas dans notre cadre dans FARGUES [ibid.].

Voici tout d'abord comment recycler le théorème de factorisation des périodes 1.3 dans un cadre joailler (ce recyclage n'était pas prévu à l'origine mais il montre combien ce théorème de factorisation des périodes est au coeur de la machine). Plus généralement que Div^1, pour $d \geq 1$, on peut définir un un faisceau pro-étale Div^d des diviseurs de Cartier effectifs de degré d sur la courbe.

Proposition 3.2 (FARGUES [ibid.]). *Le faisceau pro-étale* Div^d *est un diamant. De plus le morphisme somme de* d*-diviseurs* $\Sigma^d : (\mathrm{Div}^1)^d \to \mathrm{Div}^d$ *est quasi-pro-étale surjectif et induit un isomorphisme de faisceaux pro-étales* $(\mathrm{Div}^1)^d/\mathfrak{S}_d \xrightarrow{\sim} \mathrm{Div}^d$.

Il y a alors un morphisme d'Abel-Jacobi

$$\mathrm{AJ}^d : \mathrm{Div}^d \longrightarrow \mathcal{P}ic^d = \mathrm{Bun}^d_{\mathrm{GL}_1}.$$

On démontre alors le résultat suivant.

Théorème 3.3 (FARGUES [ibid.]). *Le morphisme* $\mathrm{AJ}^d : \mathrm{Div}^d \to \mathcal{P}ic^d$ *est une fibration pro-étale localement triviale en diamants simplement connexes si* $d > 2$.

D'après le point (2) du théorème théorème 3.1, l'application qui à un fibré en droites, fibre à fibre de degré d, associe le torseur pro-étale des isomorphismes avec $\mathcal{O}(d)$ induit une identification

$$\mathcal{P}ic^d = [\mathrm{Spa}(\overline{\mathbb{F}}_q)/\underline{E}^\times].$$

Le morphisme d'Abel-Jacobi se réécrit alors sous la forme

$$H^0(\mathcal{O}(d)) \setminus \{0\}/\underline{E}^\times \longrightarrow [\mathrm{Spa}(\overline{\mathbb{F}}_q)/\underline{E}^\times].$$

Ici $H^0(\mathcal{O}(d))$, le faisceau $S \mapsto H^0(X_S, \mathcal{O}(d))$, est un « espace de Banach-Colmez absolu » i.e. relatif au dessus de Spa$(\overline{\mathbb{F}}_q)$, l'objet final du topos pro-étale qui n'est pas représentable par un espace perfectoïde (les espaces de Banach-Colmez, qui sont des diamants, tels qu'introduits par Colmez, vivent d'habitude au dessus d'une base perfectoïde fixée, cf. Section 1.6). Bien que $H^0(\mathcal{O}(d))$ ne soit pas un diamant, on montre que $H^0(\mathcal{O}(d)) \setminus \{0\}$ en est un et est simplement connexe lorsque $d > 2$.

Références

J. Anschütz (2017). "Reductive group schemes over the Fargues-Fontaine curve". arXiv : 1703.00700 (cf. p. 324).

J. Ayoub (2015). "Motifs des variétés analytiques rigides". *Mém. Soc. Math. Fr. (N.S.)* 140-141, p. vi+386. MR : 3381140 (cf. p. 323).

L. Berger (2008a). "Construction de (ϕ, Γ)-modules : représentations p-adiques et B-paires". *Algebra Number Theory* 2.1, p. 91-120. MR : 2377364 (cf. p. 317).

– (2008b). "Équations différentielles p-adiques et (ϕ, N)-modules filtrés". *Astérisque* 319. Représentations p-adiques de groupes p-adiques. I. Représentations galoisiennes et (ϕ, Γ)-modules, p. 13-38. MR : 2493215 (cf. p. 317).

B. Bhatt, M. Morrow et P. Scholze (2015). "Integral p-adic Hodge theory – announcement". *Math. Res. Lett.* 22.6, p. 1601-1612. MR : 3507252 (cf. p. 323).

– (2016). "Integral p-adic Hodge theory". arXiv : 1602.03148 (cf. p. 323).

P Boyer (1999). "Mauvaise réduction des variétés de Drinfeld et correspondance de Langlands locale". *Invent. Math.* 138.3, p. 573-629 (cf. p. 327).

T. Bridgeland (2006). "Derived categories of coherent sheaves". In : *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, p. 563-582. MR : 2275610 (cf. p. 317).

A. Caraiani et P. Scholze (2017). "On the generic part of the cohomology of compact unitary Shimura varieties". *To appear in Annals of Math* (cf. p. 318, 324, 327, 333).

H. Cartan (1958). "Fonctions automorphes et séries de Poincaré". *J. Analyse Math.* 6, p. 169-175. MR : 0099450 (cf. p. 315).

M. Chen, L. Fargues et X. Shen (2017). "On the structure of some p-adic period domains". *Prépublication* (cf. p. 326-328).

P. Colmez (2002). "Espaces de Banach de dimension finie". *J. Inst. Math. Jussieu* 1.3, p. 331-439. MR : MR1956055 (cf. p. 316, 330).

– (2003). "Les conjectures de monodromie p-adiques". *Astérisque* 290. Séminaire Bourbaki. Vol. 2001/2002, Exp. No. 897, vii, 53-101. MR : 2074051 (cf. p. 319).

P. Colmez (2008). “Espaces vectoriels de dimension finie et représentations de de Rham”. *Astérisque* 319. Représentations p-adiques de groupes p-adiques. I. Représentations galoisiennes et (ϕ, Γ)-modules, p. 117-186. MR : MR2493217 (cf. p. 316).

P. Colmez et J.-M. Fontaine (2000). “Construction des représentations p-adiques semi-stables”. *Invent. Math.* 140.1, p. 1-43. MR : 1779803 (cf. p. 317).

P. Colmez et W. Niziol (2017). “Syntomic complexes and p-adic nearby cycles”. *Invent. Math.* 208.1, p. 1-108. MR : 3621832 (cf. p. 323).

J.-F. Dat, S. Orlik et M. Rapoport (2010). *Period domains over finite and p-adic fields*. T. 183. Cambridge Tracts in Mathematics. Cambridge : Cambridge University Press, p. xxii+372. MR : 2676072 (cf. p. 326).

V. G. Drinfeld (1983). “Two-dimensional l-adic representations of the fundamental group of a curve over a finite field and automorphic forms on $\mathrm{GL}(2)$”. *Amer. J. Math.* 105.1, p. 85-114. MR : 692107 (cf. p. 331).

G. Faltings (2002). “A relation between two moduli spaces studied by V. G. Drinfeld”. In : *Algebraic number theory and algebraic geometry*. T. 300. Contemp. Math. P. 115-129 (cf. p. 328).

– (2004). “p-adic periods”. In : *Second International Congress of Chinese Mathematicians*. T. 4. New Stud. Adv. Math. Int. Press, Somerville, MA, p. 75-80. MR : 2497972 (cf. p. 328).

L. Fargues (2004). “Cohomologie des espaces de modules de groupes p-divisibles et correspondances de Langlands locales”. In : *Variétés de Shimura, espaces de Rapoport-Zink et Correspondances de Langlands locales, Asterisque 291* (cf. p. 326).

– (2008). “L’isomorphisme entre les tours de Lubin-Tate et de Drinfeld et applications cohomologiques”. In : *L’isomorphisme entre les tours de Lubin-Tate et de Drinfeld*. Progress in math., 262. Birkhäuser, p. 1-325 (cf. p. 319, 328).

– (2010a). “Exposé à la conférence en l’honneur de Jean-Marc Fontaine” (cf. p. 317).

– (2010b). “La filtration de Harder-Narasimhan des schémas en groupes finis et plats”. *J. Reine Angew. Math.* 645, p. 1-39. MR : 2673421 (cf. p. 309, 328).

– (2011). “La filtration canonique des points de torsion des groupes p-divisibles”. *Annales scientifiques de l’ENS* 44.6, p. 905-961 (cf. p. 319).

– (2013). “Groupes rigides analytiques p-divisibles”. *Prépublication* (cf. p. 320).

– (2015a). “From local class field to the curve and vice-versa”. *À paraître dans les « Proceedings of the AMS 2015 Summer Research Institute on Algebraic Geometry » ayant eu lieu à Salt Lake city en Juillet 2015* (cf. p. 325).

– (2015b). “Lettre à Michael Rapoport” (cf. p. 327).

– (2015c). “Quelques résultats et conjectures concernant la courbe”. *Astérisque* 369, p. 325-374. MR : 3379639 (cf. p. 310, 313, 314, 316, 321, 322).

– (2016a). “G-torseurs en théorie de Hodge p-adique”. To appear in Compositio Mathematica (cf. p. 324, 325).

– (2016b). "Geometrization of the local Langlands correspondence, an overview". arXiv : 1602.00999 (cf. p. 331, 334).
– (2017). "Simple connexité des fibres d'une application d'Abel-Jacobi et corps de classe local". To appear in Annales scientifiques de l'Ecole Normale Supérieure (cf. p. 315, 316, 334).
L. Fargues et J.-M. Fontaine (2011). "Factorization of analytic functions in mixed characteristic". In : *Frontiers of mathematical sciences*. Int. Press, Somerville, MA, p. 307-315. MR : 3050829 (cf. p. 310).
– (2014). "Vector bundles on curves and p-adic Hodge theory". In : *Automorphic Forms and Galois Representations*. T. 415. London Mathematical Society Lecture Note Series. Cambridge University Press (cf. p. 310, 311).
– (2017). "Courbes et fibrés vectoriels en théorie de Hodge p-adique". To appear in Asterisque (cf. p. 309-315, 317-322).
L. Fargues et P. Scholze (p. d.). "Geometrization of the local Langlands correspondence" (cf. p. 331-334).
Laurent Fargues et Jean-Marc Fontaine (2012). "Vector bundles and p-adic Galois representations". In : *Fifth International Congress of Chinese Mathematicians. Part 1, 2*. T. 2. AMS/IP Stud. Adv. Math., 51, pt. 1. Amer. Math. Soc., Providence, RI, p. 77-113. MR : 2908062 (cf. p. 310).
J.-M. Fontaine (1979). "Modules galoisiens, modules filtrés et anneaux de Barsotti-Tate". In : *Journées de Géométrie Algébrique de Rennes. (Rennes, 1978), Vol. III*. T. 65. Astérisque. Soc. Math. France, Paris, p. 3-80. MR : 563472 (cf. p. 309).
– (1981). "Formes différentielles et modules de Tate des variétés abéliennes sur les corps locaux". *Invent. Math.* 65.3, p. 379-409 (cf. p. 319).
– (1994a). "Le corps des périodes p-adiques". *Astérisque* 223. With an appendix by Pierre Colmez, Périodes p-adiques, p. 59-111. MR : MR1293971 (cf. p. 309).
– (1994b). "Représentations p-adiques potentiellement semi-stables". In : *Périodes p-adiques*. T. 294. Astérisque, p. 321-347 (cf. p. 319).
– (2003). "Presque C_p-représentations". *Doc. Math.* Extra Vol. Kazuya Kato's fiftieth birthday, p. 285-385. MR : 2046603 (cf. p. 317).
J.-M. Fontaine et W. Messing (1987). "p-adic periods and p-adic étale cohomology". In : *Current trends in arithmetical algebraic geometry (Arcata, Calif., 1985)*. T. 67. Contemp. Math. Amer. Math. Soc., Providence, RI, p. 179-207. MR : 902593 (cf. p. 322).
I. Gaisin et N. Imai (2016). "Non-semi-stable loci in Hecke stacks and Fargues' conjecture". arXiv : 1608.07446 (cf. p. 327).
D. Gaitsgory (2002). "Geometric Langlands correspondence for GL_n". In : *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*. Higher Ed. Press, Beijing, p. 571-582. MR : 1957066 (cf. p. 331).

U. Goertz, X. He et S. Nie (2016). "Fully Hodge-Newton decomposable Shimura varieties". arXiv : 1610.05381 (cf. p. 327).

A. Grothendieck (1963). "Éléments de géométrie algébrique. III. Étude cohomologique des faisceaux cohérents. II". *Inst. Hautes Études Sci. Publ. Math.* 17, p. 91. MR : 0163911 (cf. p. 330).

D. Hansen (2016). "Moduli of local Shtukas and Harris's conjecture, I". *Prépublication* (cf. p. 327).

R. Harris et M. Taylor (2001). *The geometry and cohomology of some simple Shimura varieties*. T. 151. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ (cf. p. 327).

U. Hartl (2013). "On a conjecture of Rapoport and Zink". *Invent. Math.* 193.3, p. 627-696. MR : 3091977 (cf. p. 326).

U. Hartl et R. Pink (2004). "Vector bundles with a Frobenius structure on the punctured unit disc". *Compos. Math.* 140.3, p. 689-716. MR : MR2041777 (cf. p. 310, 314).

R. Huber (1996). *Étale cohomology of rigid analytic varieties and adic spaces.* Aspects of Mathematics. Friedr. Vieweg et Sohn, Braunschweig (cf. p. 313).

A. J. de Jong (1995). "Étale fundamental groups of non-Archimedean analytic spaces". *Compositio Math.* 97.1-2. Special issue in honour of Frans Oort, p. 89-118. MR : MR1355119 (cf. p. 326).

T. Kaletha (2016). "Rigid inner forms of real and p-adic groups". *Ann. of Math. (2)* 184.2, p. 559-632. MR : 3548533 (cf. p. 332).

K. S. Kedlaya (2004). "A p-adic local monodromy theorem". *Ann. of Math. (2)* 160.1, p. 93-184. MR : 2119719 (cf. p. 314).

– (2016a). "Noetherian properties of Fargues-Fontaine curves". *Int. Math. Res. Not. IMRN* 8, p. 2544-2567. MR : 3519123 (cf. p. 313).

– (2016b). "Some ring-theoretic properties of A_{inf}". arXiv : 1602.09016 (cf. p. 322).

K. S. Kedlaya et R. Liu (2015). "Relative p-adic Hodge theory : foundations". *Astérisque* 371, p. 239. MR : 3379653 (cf. p. 314, 315, 326, 329).

K. S. Kedlaya, J. Pottharst et L. Xiao (2014). "Cohomology of arithmetic families of (φ, Γ)-modules". *J. Amer. Math. Soc.* 27.4, p. 1043-1115. MR : 3230818 (cf. p. 313).

M. Kisin (2006). "Crystalline representations and F-crystals". In : *Algebraic geometry and number theory*. T. 253. Progr. Math. Boston, MA : Birkhäuser Boston, p. 459-496. MR : MR2263197 (cf. p. 321, 322).

R.E. Kottwitz (1985). "Isocrystals with additional structure." *Compositio Math.* 56.2, p. 201-220 (cf. p. 331).

– (1997). "Isocrystals with additional structure.II". *Compositio Math.* 109.3, p. 255-339 (cf. p. 326).

R.A. Kucharczyk et P. Scholze (2016). "Topological realizations of absolute Galois groups". arXiv : 1609.04717 (cf. p. 316).

V. LAFFORGUE (p. d.). "Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale". arXiv : 1209.5352 (cf. p. 331).

E. LAU (2016). "Dieudonné theory over semiperfect rings and perfectoid rings". arXiv : 1603.07831 (cf. p. 322).

M. LAZARD (1962). "Les zéros des fonctions analytiques d'une variable sur un corps valué complet". *Inst. Hautes Études Sci. Publ. Math.* 14, p. 47-75. MR : MR0152519 (cf. p. 311).

A.-C. LE BRAS (2017). "Espaces de Banach-Colmez et faisceaux cohérents sur la courbe de Fargues-Fontaine". To Appear in Duke Mathematical Journal (cf. p. 316).

B. LEVIN et C. WANG-ERICKSSON (2016). "A Harder-Narasimhan theory for Kisin modules". arXiv : 1606.00914 (cf. p. 328).

E. MANTOVAN (2008). "On non-basic Rapoport-Zink spaces". *Ann. Sci. Éc. Norm. Supér. (4)* 41.5, p. 671-716. MR : 2504431 (cf. p. 327).

E. MANTOVAN et E. VIEHMANN (2010). "On the Hodge-Newton filtration for p-divisible $\mathcal{O}$-modules". *Math. Z.* 266.1, p. 193-205. MR : 2670679 (cf. p. 327).

M. S. NARASIMHAN et C. S. SESHADRI (1965). "Stable and unitary vector bundles on a compact Riemann surface". *Ann. of Math. (2)* 82, p. 540-567. MR : 0184252 (cf. p. 314).

W. NIZIOL (p. d.). "Geometric syntomic cohomology and vector bundles on the Fargues-Fontaine curve". arXiv : 1605.07216 (cf. p. 323).

– (2008). "Semistable conjecture via K-theory". *Duke Math. J.* 141.1, p. 151-178. MR : 2372150 (cf. p. 322).

M. PECHE IRISSARRY (p. d.). "The reduction of G-ordinary crystalline representations with G-structure". arXiv : 1611.08529 (cf. p. 328).

M. RAPOPORT (p. d.). "Appendix A : Accessible and weakly accessible period domains, Appendix to SCHOLZE [2015]" (cf. p. 326, 327, 340).

– (1995). "Non-Archimedean period domains". In : *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*. Birkhäuser, p. 423-434 (cf. p. 331, 333).

M. RAPOPORT et E. VIEHMANN (2014). "Towards a theory of local Shimura varieties". *Münster J. Math.* 7.1, p. 273-326. MR : 3271247 (cf. p. 326).

M. RAPOPORT et T. ZINK (p. d.). *Period spaces for p-divisible groups.* Annals of Mathematics Studies 141. Princeton University Press, Princeton, NJ (cf. p. 325, 326).

P. SCHOLZE (p. d.). "Étale cohomology of diamonds". arXiv : 1709.07343 (cf. p. 315, 326, 329, 331, 332).

– (2012). "Perfectoid spaces". *Publ. Math. Inst. Hautes Études Sci.* 116, p. 245-313. MR : 3090258 (cf. p. 329).

– (2014). "Peter Scholze's lectures on p-adic geometry, Fall 2014, Berkeley" (cf. p. 315, 316, 318, 321, 322, 329, 331, 333).

P. Scholze et J. Weinstein (2013). "Moduli of p-divisible groups". *Camb. J. Math.* 1.2, p. 145-237. MR : 3272049 (cf. p. 317, 321, 328).
Peter Scholze (juin 2015). "On the p-adic cohomology of the Lubin-Tate tower". See the Appendix in Rapoport [p. d.]. arXiv : 1506.04022 (cf. p. 339).
X. Shen (2014a). "Cell decomposition of some unitary group Rapoport-Zink spaces". *Math. Ann.* 360.3-4, p. 825-899. MR : 3273647 (cf. p. 328).
– (2014b). "On the Hodge-Newton filtration for p-divisible groups with additional structures". *Int. Math. Res. Not. IMRN* 13, p. 3582-3631. MR : 3229763 (cf. p. 327).
– (2016). "On some generalized Rapoport-Zink spaces". arXiv : 1611.08977 (cf. p. 327).
T. Tsuji (1999). "p-adic étale cohomology and crystalline cohomology in the semi-stable reduction case". *Invent. Math.* 137.2, p. 233-411. MR : 1705837 (cf. p. 322).
J. Weinstein (2017). "Gal($\overline{Q}$p/Qp) as a geometric fundamental group". *Int. Math. Res. Not. IMRN* 10, p. 2964-2997. MR : 3658130 (cf. p. 316).

Received 2017-10-04.

Laurent Fargues
CNRS
Institut de Mathématiques de Jussieu
4 place Jussieu
75252 Paris
France
laurent.fargues@imj-prg.fr

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (339–362)

MULTIPLICATIVE FUNCTIONS IN SHORT INTERVALS, AND CORRELATIONS OF MULTIPLICATIVE FUNCTIONS

KAISA MATOMÄKI AND MAKSYM RADZIWIŁŁ

Abstract

Our goal in this note is two-fold. In part I, we motivate and explain the ideas behind a recent theorem of ours.

Theorem 1 (Matomäki-Radziwiłł). *Let f be a real-valued multiplicative function with $|f| \leq 1$. Then, for all $X < x \leq 2X$, with at most $o(X)$ exceptions,*

$$\frac{1}{H}\sum_{x<x\leq x+H} f(n) - \frac{1}{X}\sum_{X<n\leq 2X} f(n) = o(1)$$

as soon as $H \to \infty$ with $X \to \infty$.

In part II, which can be read independently, our goal is to survey some of the recent developments connected to Theorem 1. These have been by far and large related to Chowla's conjecture.

Conjecture 1 (Chowla). *Let μ denote the Möbius function. Then, for any set of distinct integers $h_1, \ldots, h_k$,*

$$\sum_{n\leq X} \mu(n+h_1)\ldots\mu(n+h_k) = o(X)$$

as $X \to \infty$.

Part I

We will be interested throughout in *multiplicative* functions, that is $f : \mathbb{N} \to \mathbb{C}$ such that $f(ab) = f(a)f(b)$ for all co-prime a, b. A basic example is the *Möbius function* μ, defined by $\mu(p) = -1$ and $\mu(p^\alpha) = 0$ for all $\alpha > 1$ and primes p. The Möbius function is closely connected to the primes as we now explain. Let Λ denote the von Mangoldt

MSC2010: primary 11N25; secondary 11N37.

function $\Lambda(n)$ defined by setting $\Lambda(n) = \log p$ whenever $n = p^{\alpha}$ and $\Lambda(n) = 0$ otherwise. Then, by inclusion-exclusion,

$$\Lambda(n) = -\sum_{d \mid n} \mu(d) \log d.$$

Using this relationship one can show that,

$$\sum_{n \le x} \Lambda(n) \sim x \iff \sum_{n \le x} \mu(n) = o(x) \tag{1}$$

and

$$\sum_{n \le x} \Lambda(n) = x + O_{\varepsilon}(x^{1/2+\varepsilon}) \quad \text{for all } \varepsilon > 0 \iff \sum_{n \le x} \mu(n) = O_{\varepsilon}(x^{1/2+\varepsilon}) \quad \text{for all } \varepsilon > 0. \tag{2}$$

The statement on the left-hand side of (1) is known as the Prime number theorem, while the statement on the left of (2) is an equivalent reformulation of the Riemann Hypothesis.

An immediate consequence of the Riemann Hypothesis is that the prime number theorem holds in all intervals $[x, x + x^{\alpha}]$ with $\alpha > \frac{1}{2}$. In principle deeper information than the Riemann Hypothesis is contained in the *explicit formula*,

$$\sum_{n \le x} \Lambda(n) = x - \sum_{\rho} \frac{x^{\rho}}{\rho} + O(1)$$

where the sum over ρ corresponds to a sum over zeros of the Riemann zeta-function. However from this formula it is possible to see that without understanding cancellations between zeros of the Riemann zeta-function nothing can be said about intervals with $\alpha \le \frac{1}{2}$. For $\alpha \in (0, \frac{1}{2}]$ it is thus natural to relax the question and ask whether the prime number theorem holds in most intervals $[x, x + x^{\alpha}]$ with $\alpha \in (0, \frac{1}{2}]$. Conditionally on the Riemann Hypothesis, Selberg [1943] succeeded in providing a positive answer to this question. Here we state a weak version of his result.

Theorem 2 (Selberg). *Assume the Riemann Hypothesis. Let $\varepsilon > 0$ be given. The number of integers $x \in [0, X]$ for which*

$$\Big| \sum_{x < n \le x + x^{\alpha}} \Lambda(n) - x^{\alpha} \Big| > x^{\alpha/2+\varepsilon}$$

is $o(X)$ as $X \to \infty$.

The proof of this result is not difficult : Covering the interval $[0, X]$ by dy-adic intervals, the result will follow from Chebyschev's inequality if we can show that,

$$\int_X^{2X} \Big| \sum_{x<n\le x+x^\alpha} \Lambda(n) - x^\alpha \Big|^2 dx \ll_\varepsilon X^{1+\alpha+\varepsilon}.$$

By the explicit formula, for $X \le x \le 2X$,

$$\sum_{x<n\le x+x^\alpha} \Lambda(n) - x^\alpha \approx x^\alpha \sum_{|\rho|\le X^{1-\alpha}} x^{\rho-1}.$$

where ρ is a sum over the zeros of the Riemann zeta-function, and here and later $\approx$ means that the statement is "morally true". Therefore,

$$(3) \qquad \int_X^{2X} \Big| \sum_{x<n\le x+x^\alpha} \Lambda(n) - x^\alpha \Big|^2 dx \approx X^{2\alpha} \int_X^{2X} \Big| \sum_{|\rho|\le X^{1-\alpha}} x^{\rho-1} \Big|^2 dx$$

Expanding the square, and executing the integration over $x \in [X, 2X]$ we obtain that the integral over x is

$$\approx \sum_{\substack{|\rho|\le X^{1-\alpha} \\ |\rho'|\le X^{1-\alpha}}} \frac{X^{\rho+\overline{\rho'}-1}}{\rho+\overline{\rho'}-1}$$

We then bound this trivially by

$$(4) \qquad \sum_{\substack{|\rho|\le X^{1-\alpha} \\ |\rho'|\le X^{1-\alpha}}} \frac{1}{|\rho+\overline{\rho}-1|} \ll X^{1-\alpha+\varepsilon}.$$

Altogether this shows that the left-hand side of (3) is $\ll_\varepsilon X^{1+\alpha+\varepsilon}$, as needed. A similar but more complicated argument establishes a corresponding theorem for the Möbius function.

The Riemann Hypothesis is used crucially in the upper bound (4) which uses that $\Re\rho = \Re\rho' = \frac{1}{2}$. In order to run this argument unconditionally one needs a *zero-density estimate* for the number of zeros of the Riemann zeta function in the strip $\Re s > \sigma$ and with height $|\Im s| \le T$. Using Huxley's zero-density estimate Huxley [1972] allows one to prove the following theorem (see Ramachandra [1976] for details).

Theorem 3 (Huxley). *Let $a(n) = \Lambda(n) - 1$ or $a(n) = \mu(n)$. Then, for $H > X^{1/6+\varepsilon}$, we have, for almost all $X < x < 2X$,*

$$\Big| \sum_{x<n\le x+H} a(n) \Big| \ll_A H(\log X)^{-A}.$$

Two features are worth noticing : Compared to the conditional Theorem 2, the saving that we obtain is weaker, and the range of H is worse. Crucial in the proof of Theorem 3 is the relationship of $\Lambda(n)$ or $\mu(n)$ with the zeros of the Riemann zeta-function.

Until recently this is where things stood. In a recent result we have obtained an improvement of Theorem 3 for arbitrary multiplicative function which is optimal in terms of H. We will begin with a very special case of our result for the Möbius function in the range $H = X^{\varepsilon}$.

Theorem 4 (Matomäki-Radziwiłł). *Let $\varepsilon > 0$ be given. Let $H = X^{\varepsilon}$. Then, for almost all $1 \leq x \leq X$,*

$$\sum_{x<n\leq x+H} \mu(n) = o(H).$$

Unlike Huxley's result our theorem depends directly on the multiplicativity of $\mu(\cdot)$. While the proof that we will give at first will depend on the fact that $\mu(p) = -1$ for all primes p, we will soon see that this is in no way crucial. A complete account of this special case can be found in Matomäki and Radziwiłł [2016a].

0.1 Sketch of the proof of Theorem 4. Consider

$$\int_X^{2X} \Big| \sum_{x<n\leq x+H} \mu(n) \Big|^2 dx \tag{5}$$

Our goal is to show that this is $o(XH^2)$. By an application of Plancherel, (5) is (essentially) equivalent to

$$\int_0^{X/H} \Big| \sum_{X<n\leq 2X} \frac{\mu(n)}{n^{1+it}} \Big|^2 dt = o(1). \tag{6}$$

We have at our disposal two distributional estimates for

$$\sum_{X<n\leq 2X} \frac{\mu(n)}{n^{1+it}}.$$

On the one hand the prime number theorem in the form of Vinogradov-Korobov Korobov [1958] implies that the above Dirichet polynomial is less than $O_A((\log X)^{-A})$ for all $|t| \leq \exp((\log X)^{3/2-\varepsilon})$. On the other hand a result of Montgomery and Vaughan [1974] shows that for arbitrary complex coefficients $a(n)$ and $T, X \geq 1$,

$$\int_0^T \Big| \sum_{X<n\leq 2X} a(n)n^{it} \Big|^2 dt = (T + O(X)) \sum_{X<n\leq 2X} |a(n)|^2. \tag{7}$$

Neither result is directly sufficient for obtaining (6). The first result allows us to only handle the range $|t| \le (\log X)^B$ for any fixed B. This is not sufficient to obtain the result (unless $H > X/(\log X)^B$). However, in any case it allows to reduce our attention to showing that

$$\int_{(\log X)^B}^{X/H} \Big| \sum_{X \le n \le 2X} \frac{\mu(n)}{n^{1+it}} \Big|^2 dt = o(1). \tag{8}$$

for any fixed $B > 0$. Applying (7) to (8) shows that (8) is $O(1)$. This barely fails to be non-trivial. The situation is reminiscent with what one encounters in the proof of the Bombieri-Vinogradov theorem Vaughan [1981] [1]. Similarly, the missing additional input is a bilinear structure.

We create the bilinear structure using *Ramaré's identity*,

$$\mu(n) = \sum_{\substack{n=pm \\ P \le p \le Q \\ (m,p)=1}} \mu(p) \cdot \frac{\mu(m)}{1 + \#\{P \le q \le Q : q|m\}} + \mathbf{1}_{(n, \prod_{P \le p \le Q} p)=1} \cdot \mu(n) \tag{9}$$

valid for any interval $[P, Q]$. If the parameters P, Q are chosen so that

$$\sum_{P \le p \le Q} \frac{1}{p} \to \infty$$

as $X \to \infty$, then all but $o(X)$ of the integers $X < n \le 2X$ have a prime factor in $[P, Q]$. Consequently the second term in Ramaré's identity (9) is typically zero. The point of Ramaré's identity is that it roughly allows us to write

$$\sum_{X < n \le 2X} \frac{\mu(n)}{n^{1+it}} \approx$$
$$\approx \frac{1}{\log \frac{\log Q}{\log P}} \sum_{P \le R = 2^k \le Q} \Big(\sum_{R < p \le 2R} \frac{\mu(p)}{p^{1+it}} \cdot \sum_{X/R < m \le 2X/R} \frac{\mu(m)}{m^{1+it}} \Big) + \text{“small in } L^2\text{''}.$$

[1] If one attempts to prove Bombieri-Vinogradov without a combinatorial decomposition of $\Lambda(n)$ then the available tools turn out to be insufficient: Siegel-Walfisz can handle moduli with $Q \le (\log X)^A$ while the large sieve recovers the trivial bound. This is similar to the situation we are facing here.

where "small in L^2" is a Dirichlet polynomial that is small on average. As a result, we roughly have

$$\int_{(\log X)^B}^{X/H} \Big| \sum_{X<n\le 2X} \frac{\mu(n)}{n^{1+it}} \Big|^2 dt$$
$$\text{“}\ll\text{”} \frac{\log Q}{\log R} \sum_{P\le R=2^k\le Q} \int_{(\log X)^B}^{X/H} \Big| \sum_{R<p\le 2R} \frac{1}{p^{1+it}} \cdot \sum_{X/R<m\le 2X/R} \frac{\mu(m)}{m^{1+it}} \Big|^2 dt + o(1).$$

Now we have a bilinear decomposition and we can use both distributional estimates alluded to above. First of all, if the parameter P is chosen so that $P \ge \exp((\log X)^{2/3+\varepsilon})$, then the prime number theorem of Vinogradov-Korobov applies to the Dirichlet polynomial over the primes, and gives us

$$\sum_{R<p\le 2R} \frac{1}{p^{1+it}} \ll (\log X)^{-B/2}$$

for all $R = 2^k$ in the range $[P, Q]$ and all $|t| > (\log X)^B$. Secondly, (7) shows that,

$$\int_{(\log X)^B}^{X/H} \Big| \sum_{X/R<m\le 2X/R} \frac{\mu(m)}{m^{1+it}} \Big|^2 dt \ll \Big(\frac{X}{H} + \frac{X}{R}\Big) \sum_{X/R<m\le 2X/R} \frac{1}{m^2} \ll \frac{R}{H} + 1.$$

Applying both to (8), we conclude that (8) is

$$\ll \sum_{P\le R=2^k\le Q} (\log X)^{-B} \cdot \Big(\frac{R}{H} + 1\Big) + o(1).$$

This is $o(1)$ as long $R/H \le (\log X)^{B/2}$. We collect all our requirements to see if they can be satisfied at once. We require that

1. $P \ge \exp((\log X)^{2/3+\varepsilon})$ (so that the prime number theorem is applicable).
2. $P/H \ll (\log X)^{B/2}$ (so that the mean-value theorem is efficient).
3. Q is chosen so that $\sum_{P\le p\le Q} p^{-1} \to \infty$ as $X \to \infty$ (so that in Ramaré's identity the second term is negligible).

To meet all these requirements it suffices to pick $Q = H \ge X^{\varepsilon}$ and $P = \exp((\log X)^{2/3+\varepsilon})$.

0.2 General multiplicative functions. The main thrust of the previous argument still came from a harmonious relationship between μ and prime numbers (manifested for example in the property that $\mu(p) = -1$ for all primes p). If one wishes to extend the result to general multiplicative functions, this is a bottleneck. But more generally, before we can proceed we need to understand what is the analogue of the prime number theorem for general multiplicative functions.

Since the prime number theorem can be expressed as

$$\sum_{n \le x} \mu(n) = o(x)$$

it is natural to expect that "having a prime number theorem" for $f(n)$ would amount to knowing the behavior of

$$\sum_{n \le x} f(n).$$

as $x \to \infty$. There is at present a rather well-developed theory of such mean-values. The central result is due to Halász (see also Montgomery and Vaughan [2001] and Granville, Harper, and Soundararajan [2017]).

Theorem 5 (Halász). *Let f be multiplicative with $|f| \le 1$. Then, for all $X, T \ge 1$,*

$$\frac{1}{X} \sum_{X < m \le 2X} f(m) \ll M \exp(-M) + \frac{\log \log X}{\log X} + \frac{1}{T}$$

where

$$M := \min_{|t| \le T} \sum_{p \le X} \frac{1 - \Re f(p) p^{-it}}{p}$$

The quantity M singles out cases in which $f(p)$ is close to p^{it} for most primes p. This is indeed important. If $f = m^{iT}$ then the mean-value is of size approximately X/T showing that the last term is optimal. The other two terms are unfortunately also optimal. For instance the second term takes into account multiplicative functions f such that for example $f(p) = 0$ for $p < X/2$ and $f(p) = 1$ for $X/2 \le p \le X$. In recent literature the function M is frequently denoted by $\mathbb{D}^2(f, p^{it})$ and refered to as a distance function.

Going back to our main problem, we would like to show that for almost all $X \le x \le 2X$,

$$\frac{1}{H} \sum_{x < n \le x + H} f(n) - \frac{1}{X} \sum_{X < n \le 2X} f(n) = o(1). \tag{10}$$

and for H as small as possible, compared to X. Since we aim to prove this for an arbitrary multiplicative function f with $|f| \le 1$ one should think of (10) as a statement about the factorization of the integers, rather than about multiplicative functions. Indeed, (10) states that in most short intervals $[x, x+H]$ integers factorize in a way that is similar to the long interval $[X, 2X]$. Thus in the way that integers factorize in short intervals, there can be no consistently pathological behavior that would work out to still be consistent with, for example, the prime number theorem. An example of such a pathological behavior is the existence of a positive proportion of intervals $[x, x + H]$ on which for instance $\mu(n) = 1$ and a positive proportion of intervals $[x, x + H]$ on which $\mu(n) = -1$. The intervals could be arranged in such a way so that their existence would be consistent with the prime number theorem $\sum_{n \le x} \mu(n) = o(x)$. What (10) achieves is that it rules out such a possibility.

Let us try to prove (10). Expressing both expressions in (10) in terms of Mellin transform and applying Plancherel, reveals that showing (10) amounts to proving that

$$\int_{(\log X)^{\varepsilon}}^{X/H} \Big| \sum_{X<n\le 2X} \frac{f(n)}{n^{1+it}} \Big|^2 dt = o(1).$$

Notice that the integral starts at $(\log X)^{\varepsilon}$, this is important and corresponds to the fact that the Mellin transform of the short sum and the long sum, coincide at the small frequencies $|t| \le (\log X)^{\varepsilon}$.

We now run the same argument as before. The crucial issue is to show that,

(11)
$$\int_{(\log X)^{\varepsilon}}^{X/H} \Big| \sum_{R<p\le 2R} \frac{f(p)}{p^{1+it}} \Big|^2 \cdot \Big| \sum_{X/R<m\le 2X/R} \frac{f(m)}{m^{1+it}} \Big|^2 dt = O\Big((\log X)^{-\varepsilon/2} \cdot \Big(\frac{\log Q}{\log P}\Big)^{-2}\Big)$$

for some $\varepsilon > 0$. One could get away with a smaller saving. Notice that the coefficients $f(p)$ are arbitrary, so there are no point-wise bounds for $\sum_{R<p\le 2R} f(p) p^{-1-it}$. In particular the prime number theorem is of no use!

However the sum over primes is still small for "most" t. Indeed by (7),

$$\int_{0}^{X/H} \Big| \sum_{R<p\le 2R} \frac{f(p)}{p^{1+it}} \Big|^2 dt \sim \frac{X}{H} \cdot \frac{1}{R \log R}.$$

This shows that for "most" t the sum over the primes is of size $R^{-1/2+o(1)}$. Thus for most $t \in [(\log X)^{\varepsilon}, X/H]$ our previous argument works. As a result we can focus on the set $\mathcal{U}$ of those $t \in [(\log X)^{\varepsilon}, X/H]$ for which,

$$\Big| \sum_{R<p\le 2R} \frac{1}{p^{1+it}} \Big|^2 \ge (\log X)^{-\varepsilon} \sum_{R<p\le 2R} \frac{1}{p}.$$

We can find a 1-spaced set $\mathcal{V} \subset U$ such that

$$\int_{t\in\mathcal{U}} \Big| \sum_{R<p\le 2R} \frac{f(p)}{p^{1+it}} \cdot \sum_{X/R<m\le 2X/R} \frac{f(m)}{m^{1+it}} \Big|^2 dt \ll \sum_{t\in\mathcal{V}} \Big| \sum_{R<p\le 2R} \frac{f(p)}{p^{1+it}} \Big|^2 \cdot \Big| \sum_{X/R<m\le 2X/R} \frac{f(m)}{m^{1+it}} \Big|^2 \tag{12}$$

A first task is to understand the cardinality of $\mathcal{V}$. If $Q < \exp((\log X)^{1-2\varepsilon})$ then one can compute moments

$$\int_0^{X/H} \Big| \sum_{R<p\le 2R} \frac{f(p)}{p^{1+it}} \Big|^{2k} dt$$

with a judiciously chosen k, to conclude that $|\mathcal{V}|$ is very small.

On the set $t \in \mathcal{V} \subset [(\log X)^{\varepsilon}, X/H]$ we cannot expect cancellations in the sum over the primes. In fact if there exists a single t for which there are no cancellations in both sums over p and m then we cannot succeed. Thankfully, for real-valued multiplicative functions Halász's theorem ensures that the sum over m is always non-trivially small. In fact we can hope for a saving of a small power of the logarithm in the sum over m. This leads to a bound for (12) of the form,

$$(\log X)^{-\varepsilon} \sum_{t\in\mathcal{V}} \Big| \sum_{R<p\le 2R} \frac{f(p)}{p^{1+it}} \Big|^2 \tag{13}$$

A variant of a distributional result of Halász and Turán [1969] (see Matomäki and Radziwiłł [2016b, Lemma 11]) then shows that once the set $\mathcal{V}$ is small enough, the sum over $t \in \mathcal{V}$ behaves as if there was exactly one term $t \in \mathcal{V}$ at which there are no cancellations. Therefore (13) is

$$\ll (\log X)^{-\varepsilon} \cdot \Big(\frac{1}{\log R}\Big)^2 \ll (\log X)^{-\varepsilon} \cdot (\log Q/\log P)^{-2}$$

provided that P, Q are choosen so that $P \ge \exp(\sqrt{\log Q})$, which we can assume. For instance, simply choose $Q = \exp((\log X)^{1-2\varepsilon})$ and $P = \exp((\log X)^{2/3+\varepsilon})$. This gives (11) and proves

Corollary 1 (Matomäki-Radziwiłł). *Let $f : \mathbb{N} \to [-1, 1]$ be multiplicative. Let $\varepsilon > 0$ be given and set $H = X^{\varepsilon}$. Then for almost all $X < x \le 2X$,*

$$\frac{1}{H} \sum_{x<n\le x+H} f(n) - \frac{1}{X} \sum_{X<n\le 2X} f(n) = o(1)$$

as $X \to \infty$.

0.3 The full result. In our joint work Matomäki and Radziwiłł [2016b] we succeeded in pushing the dependence on H to its optimal form. In particular we obtained the following result.

Theorem 6 (Matomäki-Radziwiłł). *Let $|f| \leq 1$ be multiplicative and real-valued. Let $\delta > 0$ be fixed. Then,*

$$\Big|\frac{1}{H}\sum_{x<n\leq x+H} f(n) - \frac{1}{X}\sum_{X<n\leq 2X} f(n)\Big| < \delta.$$

for all $X < x \leq 2X$ with $\ll X/H^{\delta/30} + X/(\log X)^{1/50}$ exceptions.

This implies that if H goes to infinity with X, no matter how slowly, then,

$$\frac{1}{H}\sum_{x<n\leq x+H} f(n) = \frac{1}{X}\sum_{X\leq n\leq 2X} f(n) + o(1) \tag{14}$$

for almost all $X \leq x \leq 2X$. This is optimal. If H were bounded then (14) is not true: for instance take $H = p+1$ with p prime and, and f equal to a quadratic character $\mod p$. Then the short sum is always equal to $1/H$ in absolute value, while the long sum is tiny. In any case when H is bounded the statement is conjectured to be false for all multiplicative functions f. However such counterexamples are open even for $f = \mu$.

It is not a matter of simple technique to go from $H = X^{\varepsilon}$ to $H \to \infty$. This can be most clearly seen in the fact that previously even on the assumption of the Riemann Hypothesis a result like (6) for $f = \mu$ was only known for $H > (\log X)^A$ for some large A.

There is in fact a very conceptual reason for this. For simplicity let's focus on the case of the Möbius function. After converting the problem to Dirichlet polynomials, for the method to work we have to create bilinear forms with Dirichlet polynomials over primes p in the range of H (or smaller, but for simplicity let's focus on p of size H). Then we have to understand the size of

$$\sum_{H\leq p\leq 2H} \frac{1}{p^{1+it}}$$

Ideally one would like to say that this Dirichlet polynomial is always small. The Riemann Hypothesis guarantees this to be the case for $H > (\log X)^{2+\varepsilon}$. When $(\log X)^{1+\varepsilon} < H < (\log X)^{2+\varepsilon}$ there is a gap in our knowledge: we still expect the Dirichlet polynomial to be always small, but even the Riemann Hypothesis is unable to confirm this. Finally the range $H < (\log X)^{1-\varepsilon}$ is particularly difficult : Using diophantine approximation one can show that there are arbitrarily large $X < t < 2X$, such that,

$$\Big|\sum_{H\leq p\leq 2H} \frac{1}{p^{1+it}}\Big| \geq \frac{1}{4}\sum_{H\leq p\leq H} \frac{1}{p}$$

Moreover the number of such t up to X is expected to be a small power of X. Therefore any argument that works in scales $H < (\log X)^{1-\varepsilon}$ will need to be able to exploit this feature to its advantage. This is not something that arguments in analytic number theory are designed to address!

The main new idea in the proof of Theorem 6 is an iterative scheme, factoring out from the Dirichlet polynomial

$$\sum_{X \le n \le 2X} \frac{f(n)}{n^{1+it}}$$

Dirichlet polynomials supported on the primes in various ranges. The argument is designed to react to the size of the Dirichlet polynomial in each range. If the Dirichlet polynomial exhibits cancellations we are done. If it does not we move to a subsequent range, but retain the information that the Dirichlet polynomial in the previous range was large. Without doing this we would not be able to succeed. For the reader interested in these details we refer to an exposition of Soundararajan [2017] or our original paper Matomäki and Radziwiłł [2016b].

Part II

While analytic number theorists have by now a coherent set of tools to tackle problems about mean-values,

$$\sum_{n \le X} a(n) \tag{15}$$

with $a(n)$ sequences of arithmetical interest, very little is known about *correlations*,

$$\sum_{n \le X} a(n)a(n+h) \tag{16}$$

with $h \neq 0$. To get a sense of the gap in the difficulty set $a(n) = \Lambda(n)$. Then (15) corresponds to the prime number theorem, while (16) is the Hardy-Littlewood 2-tuple conjecture.

If one writes,

$$\Lambda(n) = \sum_{n=ab} \mu(a) \log b \tag{17}$$

then the problem of estimating (16) reduces to that of understanding correlations of the Möbius function. For a few technical reasons we will be interested instead in correlations of the Liouville function, which differs from the Möbius function only on powers of primes. This makes in practice the two interchangeable. The Liouville function $\lambda(n)$ is defined as $\lambda(n) = (-1)^{\Omega(n)}$ where $\Omega(n)$ is the number of prime factors of n counted with multiplicity.

For our approach to succeed, we need to at the very least understand (16) with $a(n) = \mu(n)$ or $a(n) = \lambda(n)$. A conjecture of Chowla [1965] predicts that such sums always exhibit cancellations.

Conjecture 2 (Chowla)**.** *Let $a(n) = \mu(n)$ or $a(n) = \lambda(n)$. Then, for any distinct set of integers $h_1, \dots, h_\ell$,*

$$\sum_{n \le x} a(n + h_1) \dots a(n + h_\ell) = o(x).$$

as $x \to \infty$.

Instead of (17) one could use Linnik's identity,

$$\frac{\Lambda(n)}{\log n} = \sum_{k \ge 1} \frac{d_k^\star(n)}{k} \cdot (-1)^{k+1}$$

where $d_k^\star(n)$ counts the number of solutions to $n = n_1 \dots n_k$ with all $n_i > 1$. Then estimating (16) requires us to understand correlations of the kth divisor function.

Conjecture 3. *Let $h_1, \dots, h_\ell$ be a set of distinct integers and $k_1, \dots, k_\ell \ge 1$ integers. Then,*

$$\sum_{n \le x} d_{k_1}(n + h_1) \dots d_{k_\ell}(n + h_\ell) \sim C(\mathbf{k}, \mathbf{h}) \cdot x (\log x)^{k_1 + \dots + k_\ell - \ell}.$$

as $x \to \infty$, with $C(\mathbf{k}, \mathbf{h})$ a complicated constant depending on the tuples $\mathbf{k} = (k_1, \dots, k_\ell)$ and $\mathbf{h} = (h_1, \dots, h_\ell)$.

Unfortunately for us both conjectures are open. At first the second conjecture appears somewhat more approachable. For instance the problem of estimating,

$$\sum_{n} d(n) d(n+h) W\Big(\frac{n}{X}\Big) \tag{18}$$

with $W(\cdot)$ a smooth function is completely resolved. Following works of Kuznetsov we are able to write down an explicit formula for (18), with the error term involving L-functions associated to eigenfunctions of the hyperbolic Laplacian on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ ((see Motohashi [1994] for details). The triangle inequality shows that the error term is of size $O(X^{1/2+\varepsilon})$, and moreover this is optimal. One would hope that a similar strategy will work for the estimation of correlations of the third divisor function, and that the error term should involve objects related to $\mathrm{SL}_3(\mathbb{Z})$. So far such attempts have proven unsuccessful Vinogradov and Tahtadžjan [1978], and not even an asymptotic formula is known the mean-value of $d_3(n) d_3(n+1)$.

0.4 Logarithmic Chowla. However it is no longer clear that Conjecture 2 will be the first to fall. At present our understanding of such convolution sums has changed dramatically. For instance we highlight a recent result of Tao [2016b].

Theorem 7 (Tao). *We have,*

(19)
$$\sum_{n \le x} \frac{\lambda(n)\lambda(n+1)}{n} = o(\log x)$$

as $x \to \infty$.

This is refered to as *logarithmic Chowla*.

To see the relevance of Theorem 6 let us start from the very modest goal of obtaining any cancellations at all in correlations of the Liouville function. So let us suppose that

$$\sum_{n \sim X} \lambda(n)\lambda(n+1) = (1 + o(1))X$$

as $X \to \infty$, and see if we can disprove it. The above would mean that for most n, the sign of $\lambda(n)$ and $\lambda(n+1)$ is equal. Therefore there exists an H going to infinity very slowly, so that,

$$\Big| \sum_{x < n < x + H} \lambda(n) \Big| \sim H$$

in at least a positive proportion of the intervals $[x, x + H]$. Theorem 6 rules out such a possibility. In fact a quick consequence of Theorem 6 (see Matomäki and Radziwiłł [2016b, Corollary 2]) is that there exists $\delta > 0$ such that,

$$\Big| \sum_{n \le X} \lambda(n)\lambda(n+1) \Big| \le (1 - \delta)X$$

this resolved an old folklore conjecture (see for instance Hildebrand [n.d.]) and opened the door for further progress on Conjecture 1.

The next natural step is to establish Chowla's conjecture "on average". In joint work with Tao Matomäki, Radziwiłł, and Tao [2015] we obtained such a result.

Theorem 8 (Matomäki-Radziwiłł-Tao). *We have,*

(20)
$$\sum_{|h| \le H} \Big| \sum_{n \le X} \lambda(n)\lambda(n+h) \Big| = o(HX).$$

as soon as $H \to \infty$ with $X \to \infty$.

This was the crucial arithmetic ingredient in Theorem 7. Let us quickly sketch the ideas that go into the proof of Theorem 8. The identity

$$\int_0^1 \Big(\int_{\mathbb{R}} \Big| \sum_{\substack{x \le n < x+H \\ X \le n \le 2X}} \lambda(n) e(n\alpha) \Big|^2 \Big)^2 d\alpha = H \sum_{|h| \le H} (H - |h|) \Big| \sum_{n \le X} \lambda(n) \lambda(n+h) \Big|^2$$

shows that Theorem 8 will follow from

$$\sup_{\alpha \in \mathbb{R}} \sum_{X \le x \le 2X} \Big| \sum_{x \le n \le x+H} \lambda(n) e(n\alpha) \Big| = o(XH) \tag{21}$$

as $X \to \infty$ and $H \to \infty$ with X, arbitrarily slowly. This is a short-interval analogue of a classical result of Davenport (itself a variant of Vinogradov's result for primes),

$$\sup_{\alpha} \Big| \sum_{X \le n \le 2X} \lambda(n) e(n\alpha) \Big| = o(X).$$

Similarly to the proof of Davenport's theorem, the proof of (21) splits depending on the diophantine nature of α. For simplicity let us imagine that α is fixed and we are aiming at obtaining cancellations in

$$\sum_{x \le n \le x+H} \lambda(n) e(n\alpha)$$

in almost all short intervals $X \le x \le 2X$ as $X \to \infty$, and $H \to \infty$ with X. The proof splits into two cases depending on whether $\alpha \in \mathbb{Q}$ or $\alpha \notin \mathbb{Q}$. When $\alpha \notin \mathbb{Q}$ the phase $e(n\alpha)$ oscilates rather randomly, and we succeed by using ideas of Daboussi and Delange [1974], which are a variant of Vinogradov's method. This requires $f(n)$ to be multiplicative only in a certain range $[P, Q]$. [2] On the other hand when $\alpha \in \mathbb{Q}$ the phase $e(n\alpha)$ is predictable and we need to obtain cancellations from $\lambda(n)$. The most extreme case corresponds to $\alpha = 0$. However this is exactly a consequence of Theorem 6! As one might expect the proof for rational α follows by generalizing Theorem 6 to the case of arithmetic progressions. Note that in this case we use the multiplicativity of $f(n)$ in many intervals, and thus this case is significantly more arithmetic.

The estimate (21) is the crucial arithmetic input in the proof of logarithmic Chowla (19). The other input is an ingenious use of entropy allowing to replace the event $\lambda(n+p)\mathbf{1}_{p|n}$ by $\lambda(n+p)/p$, and the creation of a bilinear structure which is possible thanks to the

[2] Daboussi-Delange's work is also often refered to as the Bourgain-Katai-Sarnak-Ziegler criterion (after Bourgain, Sarnak, and Ziegler [2013], Kátai [1986]), however the correct attribution is to Daboussi-Delange who were the first to obtain such a result

logarithmic weights. In fact rather than directly attacking the Chowla conjecture, Tao shows cancellations in

$$\sum_{n \le X} \lambda(n) \Big(\sum_{\substack{p \mid n \\ P \le p \le 2P}} \lambda(n+p) \Big). \tag{22}$$

A feature of the logarithmic weights is that if one can show that (22) is non-trivially small, for any choice of P, then this implies (19). Tao's *entropy decrement argument* shows the existence of a range P on which one can replace the event $\lambda(n+p)\mathbf{1}_{p|n}$ by $\lambda(n+p)/p$. Effectively this shows that (22) is close to

$$\frac{1}{P} \sum_{P \le p \le 2P} \sum_{n \le X} \lambda(n)\lambda(n+p)$$

This is now a ternary problem, that we can hope to address by the circle method. The main ingredient here is (21): The range of P can be quite small with respect to X, and one needs a form of (21) with H of size about P.

0.4.1 Odds and ends. Of course one wonders if logarithmic Chowla has implications for prime numbers. Unfortunately while until recently this was the heuristic expectations of most experts, it turns out that the error terms in (19) are too weak to give back anything about prime numbers. This appears to be a conceptual obstruction rather than a purely technical one. The proof of (19) hinges on Theorem 6 which in turn, in the first step, discards prime numbers from consideration. Thus it seems that these methods cannot be used for a direct attack on the twin prime conjecture.

Nonetheless the logarithmic Chowla conjecture had some dramatic consequence *outside* of number theory. It led for instance to the resolution of the Erdős Discrepancy Problem Tao [2016a] in combinatorics.

Concerning the cases of logarithmic Chowla with more shifts, there has been recent progress due to Tao-Teräväinen Tao and Teräväinen [2017]. It turns out that the case of an odd number of shifts is considerably simpler, and does not require any short interval results. That this is reasonable can be perhaps most easily seen in a result of Elliott according to which, there exists a $\delta > 0$ such that,

$$\Big| \sum_{n \le x} \lambda(n)\lambda(n+1)\lambda(n+2) \Big| \le (1-\delta)X$$

The proof of the above inequality is completely elementary (see Cassaigne, Ferenczi, Mauduit, Rivat, and Sárközy [1999]). This is in stark contrast with the binary case that requires Theorem 6.

0.5 Sarnak's conjecture. One can think of Chowla's conjecture probabilistically, as asserting that if one picks a typical integer n, and we are given $\lambda(n), \lambda(n+1), \ldots, \lambda(n+k-1)$ then we get no discernible advantage in predicting $\lambda(n+k)$.

Such a point of view is appealing from the point of view of information theory: We can think of the "signal" $\lambda = (\lambda(1), \lambda(2), \lambda(3), \ldots)$, and then ask how redundant this signal is? More precisely is knowing the neighborhood of a point $\lambda(n)$ enough to reconstruct $\lambda(n)$ at least with some positive probability? If the sequence $\lambda(n)$ is truly random, then the answer should be no. In particular $\lambda(n)$ should be "orthogonal" to all sequences that have a lot of redundancies, i.e those of entropy 0.

Let us then define what we mean by the entropy of a sequence. Let $f : \mathbb{N} \to \mathbb{C}$ be an arbitrary sequence, and consider the set $S_m = \{(f(n), f(n+1), \ldots, f(n+m-1)) : n > 0\} \subset \mathbb{C}^m$ of m-tuples in $\mathbb{C}^m$. Given ε let $B(m, \varepsilon)$ be the number of m-dimensional balls in $\mathbb{C}^m$ of radius ε that are needed to cover S_m. Then, the *topological entropy of f* is defined as

$$\sigma = \sup_{\varepsilon > 0} \Big(\limsup_{m \to \infty} \frac{1}{m} \log B(m, \varepsilon) \Big).$$

Conjecture 4 (Sarnak). *Let $f : \mathbb{N} \to \mathbb{C}$, have topological entropy zero. Then,*

$$\sum_{n \leq x} \lambda(n) f(n) = o(x).$$

Sarnak's conjecture is a natural generalization of Davenport's theorem. The latter corresponds to the case of $f(n) = e(\alpha n)$ which clearly has entropy 0 (simply cover the m-dimensional unit circle by a finite union of ε balls). Sarnak's conjecture also appears naturally in additive combinatorics. A fundamental step in the proof of the Green-Tao theorem is the orthogonality of the Möbius function to nilsystems Green and Tao [2012], this corresponds to a special case of Sarnak's conjecture.

As one might expect there is a tight link between the conjectures of Chowla and Sarnak. Chowla's conjecture implies Sarnak's conjecture. Moreover if one considers the logarithmic versions of the two conjectures, then they are equivalent Tao [2017] This leads one to believe that Chowla's conjecture and Sarnak's conjecture are equivalent, but this is so far unproven.

There is a large body of literature concerning Sarnak's conjecture (see Ferenczi, Kułaga-Przymus, and Lemańczyk [2017]). There are currently two main tools : When the the sequence f is sufficiently random, one uses a criterion steeming from the work of Daboussi-Delange, also known as the Bourgain-Katai-Sarnak-Ziegler criterion. This says that if, for all fixed primes p, q,

$$\sum_{n \leq x} f(pn)\overline{f(qn)} = o(X), \tag{23}$$

then Sarnak's conjecture holds. Verifying the above condition in practice might require a substantial amount of work depending on the f under consideration. There are also cases for which (23) fails but Sarnak conjecture is expected to hold nonetheless (the simplest example being an f that changes randomly signs on blocks that grow to infinity very slowly). In such a situation one frequently tries to find a way to use Theorem 6. In fact Theorem 6 can be rephrased in purely ergodic terms.

Due to the large number of works on the subject we cannot cover it in great depth, but refer to the recent survey Ferenczi, Kułaga-Przymus, and Lemańczyk [ibid.] for the state of the art.

0.6 The shfited convolution problem on average and L-functions. When the sequence $a(n)$ is of an *automorphic* origin, then the problem of estimating the correlations,

$$\sum_{n \le x} a(n)a(n+h)$$

is often-times refered to as the *shifted convolution problem*. The reason for the different terminology is that the problem occurs frequently and naturally when trying to either estimate moments of L-functions or obtain subconvex bounds.

The prototypical moment problem is the problem of estimating,

$$M_k(T) := \int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k} dt.$$

as $T \to \infty$ where $\zeta(s)$ is the Riemann zeta-function. It is conjectured that $M_k(T) \ll T^{1+\varepsilon}$ for all $k > 0$, and even more precisely that,

$$M_k(T) = TP_k(\log T) + O(T^{1-\delta_k}) \tag{24}$$

with P_k a polynomial of degree k^2 and $\delta_k > 0$ a positive exponent. This is a powerful conjecture that implies the bound $|\zeta(\frac{1}{2}+it)| \ll_\varepsilon 1 + |t|^\varepsilon$. Such a bound (known as the Lindelöf hypothesis) would be a great substitute for the Riemann Hypothesis in many applications.

Unfortunately (24) is known only for $k = 1$ and $k = 2$. While the case of $k = 1$ can be treated with harmonic analysis alone, the estimation for $k = 2$ depends crucially on having a power-saving in the shifted convolution problem. In fact one finds that,

$$\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k} dt \approx \tag{25}$$

$$\approx TP_k(\log T) + \text{“}\frac{1}{T^{k/2-1}} \sum_{|h| \sim T^{k/2-1+\varepsilon}} e^{ih/T^{k/2-1}} \sum_{n \sim T^{k/2}} d_k(n)d_k(n+h)\text{”}$$

where the meaning of $\sim$ is left vague on purpose. Notice that the error term is trivial when $k = 1$, and is barely non-trivial for $k = 2$. Once $k \geq 3$ the error term is unfortunately too big to be manageable by any known method, and $k > 4$ seems to present extraordinary challenges as even square-root cancellation is no longer sufficient (but see B. Conrey and Keating [2015] for possible ways to circumvent this). Many results on moments focus exactly on the threshold where the sum in the error term is barely non-trivial.

0.6.1 The shifted convolution problem for $d_k(n)$. The shape of the error term in (25) motivates us to understand,

$$\sum_{n \leq X} d_k(n) d_k(n+h)$$

on average over h, or more generally with $d_k(n)$ replaced by coefficients of a $GL(k)$ automorphic form. The literature contains average results when the number of shifts $|h| \leq H$ is substantial, i.e $H > X^{1/3+\varepsilon}$ (and recently for $H > X^{8/33+\varepsilon}$). However using methods related to Theorem 6 one can reduce the number of shifts to be as small as $(\log X)^{O(k \log k)}$.

Theorem 9 (Matomäki-Radziwiłł-Tao)**.** *Let $k \geq \ell \geq 2$ be real numbers. Then, for $H = (\log X)^{10000 k \log k}$,*

$$\sum_{|h| \leq H} \Big| \sum_{n \leq X} d_k(n) d_\ell(n+h) - X P_{k+\ell}(\log X) \Big| = o(HX (\log X)^{k+\ell-2})$$

where $P_{k+\ell}$ is a polynomial of degree $k + \ell - 2$.

Compared to our earlier work multiplicative functions such as $d_k(n)$ present new challenges because they are unbounded. But it is not the fact that they are unbounded in itself which is the main difficulty – it's rather the fact that because they are unbounded, the main contribution to

$$\sum_{n \leq X} d_k(n) d_k(n+h)$$

comes from a thin subset of integers having $(k + o(1)) \log \log x$ prime factors and on which $d_k(n)$ is unusually large. The density of such integers is about $(\log X)^{-k \log k + k - 1}$.

The idea behind the proof of Theorem 9 is to first restrict to a density one subset of integers $\mathfrak{N}$ on which we can construct efficient sieve majorants for $d_k(n)$. Subsequently after some harmonic analysis, we find that the crucial issue is to obtain a non-trivial estimate for

$$\sum_{j=1}^{J} \sum_{X \leq x \leq 2X} \Big| \sum_{\substack{x < n < x+H \\ n \in \mathfrak{N}}} d_k(n) e(n \alpha_j) \Big|$$

where α_j is a set of well-spaced points and J is arbitrary. When J is large one can simply appeal to the large sieve. On the other hand when J is bounded obtaining a non-trivial result amounts to a variant of Theorem 6 for unbounded multiplicative functions (similar variants for other sparse sets of integers were studied in Goudout [2017] and Teräväinen [2016]). It is the intermediate range between bounded and large J which turns out the most subtle. In this range through a use of duality we (essentially) replace $d_k(n)$ by the corresponding sieve majorant $\widetilde{d}_k(n)$, and we are reduced to needing cancellations in

$$\sum_{x<n<x+H} \widetilde{d}_k(n)e(n\alpha_j)$$

after a substantial amount of effort (this would be trivial if $H > X^{\varepsilon}$). We then estimate high moments of such a sum to conclude that it exhibits cancellations for most x.

In fact the argument works quite generally for estimating $d_k(n)b(n+h)$ with $b(n)$ any sequence for which efficient sieve majorants can be constructed. For instance, we obtain results for the higher-order Titchmarsh divisor problem,

$$\sum_{n\leq X} d_k(n)\Lambda(n+h)$$

with an average over $|h| \leq H$ of size at most $H = (\log X)^{10000k\log k}$. For individual h the problem is open for all $k > 2$.

0.6.2 Moments of L-functions. Coming back to the moment problem, our results (such as Theorem 9) say nothing new in the case of the Riemann zeta-function, but they are nonetheless useful in other families. For instance, a q-analogue of moments of the Riemann zeta-function, is the problem of estimating,

$$\sum_{q\leq Q}\sum_{\chi \pmod q}\int_{\mathbb{R}} |L(\tfrac{1}{2}+it,\chi)|^{2k}\cdot\Phi(t)dt \tag{26}$$

where χ is a sum over primitive characters, $L(s,\chi)$ is a Dirichlet L-function and $\Phi(t)$ is a fixed smooth function. This problem (but without the t averaging) is also related to understanding the distribution of $d_k(n)$ in arithmetic progressions.

The large sieve gives sharp upper bound for the above moment problem when $k \leq 4$. A few years ago, J. B. Conrey, Iwaniec, and Soundararajan [2013] devised a method to obtain asymptotic estimates in moments such as (26). They illustrated their method to obtain an asymptotic for (26) when $k = 3$. The case $k = 4$ represents the absolute limit of their method, and also the limit of what should be realistically feasible. In Chandee and Li [2014] it was addressed conditionally on the Generalized Riemann Hypothesis.

Roughly for the solution of the case $k = 4$ one needs to estimate on average a shifted convolution problem in a short interval. It turns out that this problem is amenable to our earlier methods and leads to the following Theorem (the write-up is currently in preparation).

Theorem 10 (Chandee-Li-Matomäki-Radziwiłł). *Let $\Phi(t)$ be a fixed smooth function. Then,*

$$\sum_{q \leq Q} \sum_{\chi \pmod q} \int_{\mathbb{R}} |L(\tfrac{1}{2} + it, \chi)|^8 \cdot \Phi(t) dt \sim C \widehat{\Phi}(0) \cdot Q^2 (\log Q)^{16}.$$

as $Q \to \infty$, where the summation is over primitive characters and $C > 0$ is an absolute constant.

Previously the assumption of the Generalized Riemann Hypothesis was used to estimate non-trivially the shifted convolution problem that arises in this problem.

0.6.3 Gaps between multiplicative sequences. The applications of the shifted convolution problem are not restricted to problems related in one way or another to L-functions. A prominent example is Hooley's Hooley [1971] work on gaps between sums of two squares. Let $1 = s_1 < s_2 < \ldots$ be the sequence of integers representable as sums of two squares. Then the average gap between $s_{n+1} - s_n$ for $s_n \leq x$ is $\asymp \sqrt{\log x}$. Hooley investigated how often the gaps deviate from the mean. He proved that for $\gamma < 5/3$,

$$\sum_{s_n \leq x} (s_{n+1} - s_n)^{\gamma} \asymp x (\log x)^{\frac{1}{2}(\gamma - 1)}. \tag{27}$$

The form of the expression is motivated by Erdős's conjecture that,

$$\sum_{p_n \leq x} (p_{n+1} - p_n)^2 \asymp x \log x.$$

In (27) the lower bound is easy and follows from Holder's inequality against the case $\gamma = 1$. In principle (27) is conjectured to hold for all finite γ, but this is a very deep conjecture. It implies for instance that for any fixed $\varepsilon > 0$ in all intervals of the form $[x, x + x^{\varepsilon}]$ there is a sum of two squares.

The problem of showing (27) is equivalent to obtaining the following frequency bound,

$$\#\Big\{x \leq X : \sum_{x < s_n \leq x + H\sqrt{\log X}} 1 = 0\Big\} \ll \frac{X}{H^{\gamma - 1}} \tag{28}$$

uniformly in $1 \leq H \leq X$. It is the uniformity that is difficult to maintain in this problem.

Hooley's approach goes as follows: In order to obtain (28) he notices that for any weight w_n supported on sums of two squares, and for any constant A, the frequency (28) is

$$\ll \frac{1}{A^2} \sum_{x \leq X} \Big| \sum_{x \leq n \leq x+H} w_n - A \Big|^2 \tag{29}$$

Moreover the estimation of (29) is feasible if we can get a good estimate for the correlations

$$\sum_{|h| \leq H} \sum_{n \leq X} w_n \overline{w_{n+h}}$$

that saves at least a power of H over the trivial bound.

When H is small, Hooley selects $w_n = r(n)\rho(n)$ where $r(n)$ is the number of representations of n as a sums of two squares, and $\rho(n)$ is a sieve weight dampening the size of $r(n)$ so that $r(n)\rho(n) \approx 1$ on average. When H is large (say $H > X^{\varepsilon}$) we can afford to loose powers of the logarithm, and it's enough to choose $w_n = r(n)$. Then we require a shifted convolution on average, with a power-saving. This is possible to obtain because $r(n)$ is roughly similar to the divisor function, and similar techniques that can be used to estimate,

$$\sum_{n \leq x} d(n)d(n+h)$$

will also do the job for the $r(n)$ function.

The sequence of sums of two squares is a norm-form, since $a^2 + b^2$ is the norm of Gaussian integers. It is natural to ask if Hooley's result can be extended to norm forms of higher degree fields. Unfortunately we run right away in a serious difficulty : If K is a number field of degree 3 and $r_K(n)$ is the number of representations of n as a norm of an ideal in K, then there are no results with power-savings for the shifted convolution problem,

$$\sum_{|h| \leq X^{\delta}} \sum_{n \leq X} r_K(n) r_K(n+h)$$

In fact this is of a comparable difficulty as our Conjecture 2 in the case $\ell = 2$ and $k_1 = k_2 = 3$.

It turns out however that we can circumvent these difficulties by using techniques related to Theorem 6. The advantage of using these "more restrictive techniques" (after all we forfeit any possibility of power-savings) is that not only we can extend the result to norm-forms, but more generally to any "multiplicative sequence" (of which norm-forms are an example).

Corollary 2 (Matomäki-Radziwiłł). *Let $\mathcal{P}$ be a set of primes of positive density δ. Let $\mathcal{N}$ be the set of all square-free integers all of whose prime factors belong to $\mathcal{P}$. Denote the elements of $\mathcal{N}$ by $n_1 < n_2 < \dots$. Then, for all $\gamma < \frac{3}{2}$,*

$$\sum_{n_k \le x} (n_{k+1} - n_k)^{\gamma} \asymp_{\mathcal{P},\gamma} x(\log x)^{(1-\delta)(\gamma-1)}. \tag{30}$$

We think it is remarkable that the exponent $\frac{3}{2}$ does not depend on the density of $\mathcal{P}$. Let us very quickly explain the kind of ideas that go into the proof of Corollary (2). The case of H small can be disposed by proving a variant of our Theorem 6 for multiplicative functions that are supported on a set of primes of density $0 < \delta < 1$.

When H is large, the techniques that go into Theorem 6 are not immediately applicable, and need to be modified. Let us highlight the spirit of these modifications in the case of norm forms. What makes the shifted convolution problem (30) difficult are certain specific sets of integers, for instance the integers n that factor into abc with a, b, c roughly of equal size. However if one restricts in the shifted convolution problem to integers of the form $n = abc$ with a, b, c in a certain special configuration then the problem can be solved with a power-saving in X. Unfortunately the set of integers in such a desirable configuration might be of density zero, but this is not a problem when H is large!

The reason is that in the regime H large we can afford to lose some powers of the logarithm, and by restricting to a density zero subset (instead of the full sequence) we are typically loosing at most powers of the logarithm. So we simply run Hooley's method on the subsequence of integers representable as norm forms and that factor as $n = abc$ with abc in certain desirable configurations. In reality the proof of Corollary 2 goes along completely different lines, however what we explained highlights the spirit of the ideas.

0.7 The structure of multiplicative functions.

Our Theorem 6 is of course rather immediately applicable to the study of general multiplicative functions. For instance it immediately implies the following result.

Corollary 3. *Let $f : \mathbb{N} \to \mathbb{R}$ be a multiplicative function which is non-zero for positive proportion of natural numbers. Then f has a positive proportion of sign changes if and only if $f(n) < 0$ for some $n \in \mathbb{N}$.*

This was an improvement even in the case of $f(n) = \mu(n)$. This result has been the starting point of several "rigidity theorems" of Klurman [2017] and Klurman and Mangerel [2017]. We highlight a few of their results. For instance, in Klurman [2017] an old conjecture of Katai is resolved. This asserts that for f taking values on the unit disk,

$$\sum_{n \le x} |f(n+1) - f(n)| = o(x) \tag{31}$$

if and only if, $f(n) = n^{it}$ for some $t \in \mathbb{R}$, or

$$\sum_{n \leq x} |f(n)| = o(x).$$

Notice that if f is assumed to be real-valued and lying on the unit-disk (i.e $f(n) = \pm 1$) then Katai's conjecture is nothing more than a convoluted restatement of Corollary 3.

Moreover in Klurman and Mangerel [2017] appears a solution to an old conjecture of Chudakov. The conjecture states that if f is a multiplicative function such that $f(n)$ takes only finitely many values, and $f(p)$ is zero on only finitely many primes, and

$$\sum_{n \leq x} f(n) = \alpha x + O(1)$$

then f is a Dirichlet character. In a similar vein one can think of the Erdős Discrepancy Problem as a rigidity theorem, since it implies that there are no completely multiplicative functions f with $\sum_{n \leq x} f(n) = O(1)$ for all x.

References

J. Bourgain, P. Sarnak, and T. Ziegler (2013). "Disjointness of Moebius from horocycle flows". In: *From Fourier analysis and number theory to Radon transforms and geometry*. Vol. 28. Dev. Math. Springer, New York, pp. 67–83. MR: 2986954 (cit. on p. 352).

Julien Cassaigne, Sébastien Ferenczi, Christian Mauduit, Jöel Rivat, and András Sárközy (1999). "On finite pseudorandom binary sequences. III. The Liouville function. I". *Acta Arith.* 87.4, pp. 367–390. MR: 1671629 (cit. on p. 353).

Vorrapan Chandee and Xiannan Li (2014). "The eighth moment of Dirichlet L-functions". *Adv. Math.* 259, pp. 339–375. MR: 3197660 (cit. on p. 357).

S. Chowla (1965). *The Riemann hypothesis and Hilbert's tenth problem*. Mathematics and Its Applications, Vol. 4. Gordon and Breach Science Publishers, New York-London-Paris, pp. xv+119. MR: 0177943 (cit. on p. 350).

Brian Conrey and Jonathan P. Keating (2015). "Moments of zeta and correlations of divisor-sums: I". *Philos. Trans. Roy. Soc. A* 373.2040, pp. 20140313, 11. MR: 3338122 (cit. on p. 356).

J. Brian Conrey, Henryk Iwaniec, and Kannan Soundararajan (2013). "Critical zeros of Dirichlet L-functions". *J. Reine Angew. Math.* 681, pp. 175–198. MR: 3181494 (cit. on p. 357).

Hédi Daboussi and Hubert Delange (1974). "Quelques propriétés des fonctions multiplicatives de module au plus égal à 1". *C. R. Acad. Sci. Paris Sér. A* 278, pp. 657–660. MR: 0332702 (cit. on p. 352).

S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk (Oct. 2017). "Sarnak's Conjecture – what's new". arXiv: 1710.04039 (cit. on pp. 354, 355).

Élie Goudout (2017). "Lois locales de la fonction ω dans presque tous les petits intervalles". *Proc. Lond. Math. Soc. (3)* 115.3, pp. 599–637. MR: 3694294 (cit. on p. 357).

Andrew Granville, Adam J Harper, and K. Soundararajan (June 2017). "A new proof of Halász's Theorem, and its consequences". arXiv: 1706.03749 (cit. on p. 345).

Ben Green and Terence Tao (2012). "The Möbius function is strongly orthogonal to nilsequences". *Ann. of Math. (2)* 175.2, pp. 541–566. MR: 2877066 (cit. on p. 354).

G. Halász and P. Turán (1969). "On the distribution of roots of Riemann zeta and allied functions. I". *J. Number Theory* 1, pp. 121–137. MR: 0236124 (cit. on p. 347).

Adolf Hildebrand (n.d.). review. no. 95d:11099 (cit. on p. 351).

Christopher Hooley (1971). "On the intervals between numbers that are sums of two squares". *Acta Math.* 127, pp. 279–297. MR: 0294281 (cit. on p. 358).

M. N. Huxley (1972). "On the difference between consecutive primes". *Invent. Math.* 15, pp. 164–170. MR: 0292774 (cit. on p. 341).

I. Kátai (1986). "A remark on a theorem of H. Daboussi". *Acta Math. Hungar.* 47.1-2, pp. 223–225. MR: 836415 (cit. on p. 352).

Oleksiy Klurman (2017). "Correlations of multiplicative functions and applications". *Compos. Math.* 153.8, pp. 1622–1657. MR: 3705270 (cit. on p. 360).

Oleksiy Klurman and Alexander P. Mangerel (July 2017). "Rigidity Theorems for Multiplicative Functions". arXiv: 1707.07817 (cit. on pp. 360, 361).

N. M. Korobov (1958). "Estimates of trigonometric sums and their applications". *Uspehi Mat. Nauk* 13.4 (82), pp. 185–192. MR: 0106205 (cit. on p. 342).

Kaisa Matomäki and Maksym Radziwiłł (2016a). "Multiplicative functions in short intervals". *Ann. of Math. (2)* 183.3, pp. 1015–1056. arXiv: 1502.02374. MR: 3488742 (cit. on p. 342).

– (2016b). "Multiplicative functions in short intervals". *Ann. of Math. (2)* 183.3, pp. 1015–1056. MR: 3488742 (cit. on pp. 347–349, 351).

Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao (2015). "An averaged form of Chowla's conjecture". *Algebra Number Theory* 9.9, pp. 2167–2196. MR: 3435814 (cit. on p. 351).

H. L. Montgomery and R. C. Vaughan (1974). "Hilbert's inequality". *J. London Math. Soc. (2)* 8, pp. 73–82. MR: 0337775 (cit. on p. 342).

– (2001). "Mean values of multiplicative functions". *Period. Math. Hungar.* 43.1-2, pp. 199–214. MR: 1830577 (cit. on p. 345).

Yōichi Motohashi (1994). "The binary additive divisor problem". *Ann. Sci. École Norm. Sup. (4)* 27.5, pp. 529–572. MR: 1296556 (cit. on p. 350).

K. Ramachandra (1976). "Some problems of analytic number theory". *Acta Arith.* 31.4, pp. 313–324. MR: 0424723 (cit. on p. 341).

Atle Selberg (1943). "On the normal density of primes in small intervals, and the difference between consecutive primes". *Arch. Math. Naturvid.* 47.6, pp. 87–105. MR: 0012624 (cit. on p. 340).

Kannan Soundararajan (2017). "The Liouville function in short intervals". *Astérisque* 390. Séminaire Bourbaki. Vol. 2015/2016. Exposés 1104–1119, Exp. No. 1119, 453–479. MR: 3666035 (cit. on p. 349).

Terence Tao (2016a). "The Erdős discrepancy problem". *Discrete Anal.* Paper No. 1, 29. MR: 3533300 (cit. on p. 353).

– (2016b). "The logarithmically averaged Chowla and Elliott conjectures for two-point correlations". *Forum Math. Pi* 4, e8, 36. MR: 3569059 (cit. on p. 351).

– (2017). "Equivalence of the logarithmically averaged Chowla and Sarnak conjectures". In: *Number theory—Diophantine problems, uniform distribution and applications*. Springer, Cham, pp. 391–421. MR: 3676413 (cit. on p. 354).

Terence Tao and Joni Teräväinen (Aug. 2017). "The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures". arXiv: 1708.02610 (cit. on p. 353).

Joni Teräväinen (2016). "Almost primes in almost all short intervals". *Math. Proc. Cambridge Philos. Soc.* 161.2, pp. 247–281. MR: 3530507 (cit. on p. 357).

R. C. Vaughan (1981). "An elementary method in prime number theory". In: *Recent progress in analytic number theory, Vol. 1 (Durham, 1979)*. Academic Press, London-New York, pp. 341–348. MR: 637357 (cit. on p. 343).

A. I. Vinogradov and L. A. Tahtadžjan (1978). "Theory of the Eisenstein series for the group $SL(3, \mathbf{R})$ and its application to a binary problem. I. Fourier expansion of the highest Eisenstein series". *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* 76. Analytic number theory and the theory of functions, pp. 5–52, 216. MR: 527787 (cit. on p. 350).

Received 2017-12-09.

Kaisa Matomäki
ksmato@utu.fi

Maksym Radziwiłł
maksym.radziwill@mcgill.ca

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (363–380)

GAPS BETWEEN PRIMES

James Maynard

Abstract

We discuss recent advances on weak forms of the Prime k-tuple Conjecture, and its role in proving new estimates for the existence of small gaps between primes and the existence of large gaps between primes.

1 Introduction

It follows from the Prime Number Theorem that the *average* gap between primes less than X is of size roughly $\log X$ when X is large. We expect, however, that occasionally these gaps are rather smaller than $\log X$, and occasionally rather larger. Specifically, based on random models and numerical evidence, we believe that the largest and smallest gaps are as described in the following two famous conjectures[1].

Conjecture 1 (Twin Prime Conjecture)**.** *There are infinitely many pairs of primes which differ by exactly 2.*

Conjecture 2 (Cramér's Conjecture, weak form)**.** *Let p_n denote the n^{th} prime. Then*

$$\sup_{p_n \le X} (p_{n+1} - p_n) = (\log X)^{2+o(1)}.$$

Moreover, the Twin Prime Conjecture can be thought of as a special case of the far-reaching Prime k-tuple Conjecture describing more general patterns of many primes.

Conjecture 3 (Prime k-tuple Conjecture)**.** *Let $L_1, \dots, L_k$ be integral linear functions $L_i(n) = a_i n + b_i$ such that for every prime p there is an integer n_p with $\prod_{i=1}^k L_i(n_p)$ coprime to p. Then there are infinitely many integers n such that all of $L_1(n), \dots, L_k(n)$ are primes.*

The author is funded by a Clay Research Fellowship. This article was written whilst the author was a member of the Institute for Advanced Study, Princeton and supported by the National Science Foundation under Grant No. DMS - 1638352.

MSC2010: primary 11N05; secondary 11N35.

[1] The original conjecture of Cramér [1936] was a stronger statement which is no longer fully believed to be true. We expect the weaker version given here to hold.

Unfortunately all these conjectures seem well beyond the current techniques, but we are able to make partial progress by showing that we do have gaps which are smaller or larger than the average gap. A key part of recent progress on results about gaps between primes has been weak versions of Conjecture 3, where one shows there are infinitely many integers n such that *several* (rather *all*) of the linear functions are prime at n.

Theorem 1 (Maynard [2015]). *Let $L_1, \dots, L_k$ be integral linear functions $L_i(n) = a_i n + b_i$ such that for every prime p there is an integer n_p with $\prod_{i=1}^k L_i(n_p)$ coprime to p. Then there is a constant $c > 0$ such that there are infinitely many integers n where at least $c \log k$ of $L_1(n), \dots, L_k(n)$ are primes.*

Variants of Theorem 1 have been important in recent results on small and large gaps between primes, and have proven useful because the method of proof is quite flexible and can generalize to other situations. Specifically, we now know the following approximations to Conjecture 1 and Conjecture 2.

Theorem 2 (Polymath [2014]). *There are infinitely many pairs of primes which differ by at most* 246.

Theorem 3 (Ford, Green, Konyagin, Maynard, and Tao [2018]). *There exists a constant $c > 0$ such that*

$$\sup_{p_n \le X} (p_{n+1} - p_n) \ge c \frac{\log X \cdot \log\log X \cdot \log\log\log\log X}{\log\log\log X}.$$

Moreover, in the case of small gaps between primes, we can show the existence of *many* primes in bounded length intervals.

Theorem 4 (Maynard [2016a]). *There exists a constant $c > 0$ such that, for all $X \ge 2$ there are at least $cX \exp(-\sqrt{\log X})$ integers $x \in [X, 2X]$ such that*

$$\#\{\textit{primes in } [x, x+y]\} \ge c \log y.$$

For example, this shows that there are infinitely many intervals of length $e^{m/c}$ containing m primes, and, for fixed $\epsilon > 0$, infinitely many x such that the interval $[x, x+(\log x)^\epsilon]$ contains $\epsilon c \log\log x$ primes.

2 The GPY sieve method

We aim to prove Theorem 1 by the 'GPY method', which can be interpreted as a first moment method which was introduced to study small gaps between primes by Goldston, Pintz, and Yıldırım [2009]. This takes the following basic steps, for some given set $\{L_1, \dots, L_k\}$ of distinct functions satisfying the hypotheses of Conjecture 3:

1. We choose a probability measure w supported on integers in $[X, 2X]$.

2. We calculate the expected number of the functions $L_i(n)$ which are prime at n, if n is chosen randomly with probability $w(n)$.

3. If this expectation is at least m, then there must be some $n \in [X, 2X]$ such that at least m of $L_1(n), \dots, L_k(n)$ are primes.

4. If this holds for all large X, then there are infinitely many such integers n.

This procedure only works if we can find a probability measure w which is suitably concentrated on integers n where many of the $L_i(n)$ are prime, but at the same time is simple enough that we can calculate this expectation unconditionally. We note that by linearity of expectation, it suffices to be able to calculate the probability that any one of the linear functions is prime. However, any slowly changing smooth function w is insufficient since the primes have density 0 in the integers, whereas any choice of w explicitly depending on the joint distribution of prime values of the L_i is likely to be too complicated to handle unconditionally.

Sieve methods are a flexible set of tools, developed over the past century, which provide natural choices for the probability measure w. The situation of simultaneous prime values of $L_1, \dots, L_k$ is a 'k-dimensional' sieve problem. For such problems when k is large but fixed, the Selberg sieve tends to be the type of sieve which performs best. The standard choice of Selberg sieve weights (which are essentially optimal in closely related situations) are

$$w(n) \propto \Bigg(\sum_{\substack{d \mid \prod_{i=1}^k L_i(n) \\ d<R}} \mu(d) \Big(\log \frac{R}{d}\Big)^k \Bigg)^2, \tag{2-1}$$

where R is a parameter which controls the complexity of the sieve weights, and w is normalized to sum to 1 on $[X, 2X]$.

To calculate the probability that $L_j(n)$ is prime with this choice of $w(n)$, we wish to estimate the sum of $w(n)$ over $n \in [X, 2X]$ such that $L_j(n)$ is prime. To do this we typically expand the divisor sum in the definition (2-1) and swap the order of summation. This reduces the problem to estimating the number of prime values of $L_j(n)$ for $n \in [X, 2X]$ in many different arithmetic progressions with moduli of size about R^2. The Elliott-Halberstam conjecture Elliott and Halberstam [1970] asserts that we should be able to do this when $R^2 < X^{1-\epsilon}$, but unconditionally we only know how to do this when $R^2 < X^{1/2-\epsilon}$, using the Bombieri-Vinogradov Theorem Bombieri [1965] and Vinogradov [1965]. After some computation one finds that, provided we do have suitable estimates for primes in arithmetic progressions, the choice (2-1) gives

$$\mathbb{E}\,\#\{1 \le i \le k : L_i(n) \text{ prime}\} = \Big(2 - \frac{2}{k+1} + o(1)\Big)\frac{\log R}{\log X}. \tag{2-2}$$

In particular, this is less than 1 for all large X, even if we optimistically assume the Elliott-Halberstam conjecture and take $R \approx X^{1/2-\epsilon}$, and so it appears that we will not be able to conclude anything about primes in this manner.

The groundbreaking work of Goldston, Pintz, and Yıldırım [2009] showed that a variant of this choice of weight actually performs much better. They considered

$$w(n) \propto \Big(\sum_{\substack{d \mid \prod_{i=1}^k L_i(n) \\ d<R}} \mu(d)\Big(\log\frac{R}{d}\Big)^{k+\ell}\Big)^2, \tag{2-3}$$

where ℓ is an additional parameter to be optimized over. With the choice (2-3), one finds that provided we have suitable estimates for primes in arithmetic progressions, we obtain

$$\mathbb{E}\,\#\{1 \le i \le k : L_i(n) \text{ prime}\} = \Big(4 + O\Big(\frac{1}{\ell}\Big) + O\Big(\frac{\ell}{k}\Big)\Big)\frac{\log R}{\log X}.$$

This improves upon (2-2) by a factor of about 2 when k is large and $\ell \approx k^{1/2}$. This falls just short of proving that two of the linear functions are simultaneously prime when $R = X^{1/4-\epsilon}$ as allowed by the Bombieri-Vinogradov theorem, but any small improvement allowing $R = X^{1/4+\epsilon}$ would give bounded gaps between primes! By considering additional possible primes, Goldston Pintz and Yıldırım were able to show

$$\liminf_n \frac{p_{n+1} - p_n}{\log p_n} = 0,$$

finally extending a long sequence of improvements to upper bounds for the left hand side Erdős [1940], Rankin [1950], Ricci [1954], Bombieri and Davenport [1966], Pilt'ja [1972], Uchiyama [1975], M. N. Huxley [1973, 1977], M. Huxley [1984], and Maier [1988].

Building on earlier work of Fouvry and Iwaniec [1980] and Bombieri, Friedlander, and Iwaniec [1986, 1987, 1989], Zhang [2014] succeeded in establishing an extended version of the Bombieri-Vinogradov Theorem for moduli with no large prime factors allowing $R = X^{1/4+\epsilon}$, and ultimately this allowed him to show

$$\mathbb{E}\,\#\{1 \le i \le k : L_i(n) \text{ prime}\} > 1,$$

if $k > 3\,500\,000$ and X is sufficiently large. The key breakthrough in Zhang's work was this result on primes in arithmetic progressions of modulus slightly larger than $X^{1/2}$. By choosing suitable linear functions, this then showed that

$$\liminf_n (p_{n+1} - p_n) \leq 7 \cdot 10^7.$$

3 A modified GPY sieve method

An alternative approach to extending the work of Goldston, Pintz and Yıldırım was developed independently by the author Maynard [2015] and Tao (unpublished). The key difference was to consider the multidimensional generalization

$$w(n) \propto \Big(\sum_{\substack{d_1,\dots,d_k \\ d_i | L_i(n) \\ \prod_{i=1}^k d_i < R}} \mu(d) F\Big(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R}\Big)\Big)^2, \tag{3-1}$$

for suitable smooth functions $F : \mathbb{R}^k \to \mathbb{R}$ supported on $[0,\infty)^k$. The flexibility of allowing the function F to depend on each divisor $d_1, \dots, d_k$ of $L_1(n), \dots, L_k(n)$ allows us to make $w(n)$ more concentrated on integers n when many of the $L_i(n)$ are prime.

With this choice, after some computation, one finds that

$$\mathbb{E}\,\#\{1 \leq i \leq k : L_i(n) \text{ prime}\} = \Big(\frac{\sum_{i=1}^k J_i(\tilde{F})}{I(\tilde{F})} + o(1)\Big)\frac{\log R}{\log X}, \tag{3-2}$$

provided, as before, we are able to count primes in arithmetic progressions to modulus R^2 on average. Here the $J_i(\tilde{F})$ and $I(\tilde{F})$ are k-dimensional integrals depending on a transform[2] $\tilde{F}$ of F, given by

$$J_\ell(\tilde{F}) = \int \cdots \int_{\sum_{i\neq\ell} t_i \leq 1} \Big(\int_0^{1-\sum_{i\neq\ell} t_i} \tilde{F}(t_1, \dots, t_k) dt_\ell \Big)^2 dt_1 \dots dt_{\ell-1} dt_{\ell+1} \dots dt_k,$$

$$I(\tilde{F}) = \int \cdots \int_{\sum_{i=1}^k t_i \leq 1} \tilde{F}(t_1, \dots, t_k)^2 dt_1 \dots dt_k.$$

For any piecewise smooth choice of $\tilde{F}$ supported on $\sum_{i=1}^k t_i \leq 1$ there is a corresponding choice of F. In particular, we can show that many of the L_i are simultaneously prime infinitely often, if we can show that $\sup_{\tilde{F}} (\sum_{i=1}^k J_i(\tilde{F})/I(\tilde{F})) \to \infty$ as $k \to \infty$.

[2] $\tilde{F}$ is F differentiated with respect to each coordinate.

A key advantage of this generalization is that we can make use of high dimensional phenomena such as concentration of measure. We concentrate on functions $\tilde{F}$ the form

$$\tilde{F}(t_1,\dots,t_k)=\begin{cases}\prod_{i=1}^k G(t_i), & \sum_{i=1}^k t_i<1,\\ 0, & \text{otherwise},\end{cases}$$

and make a probabilistic interpretation of the integrals $J_i(\tilde{F})$ and $I(\tilde{F})$. Let $Z_1,\dots,Z_k$ be i.i.d. random variables on $[0,\infty]$ with probability density function G^2, expectation $\mu=\int_0^\infty tG(t)^2dt$ and variance $\sigma^2=\int_0^\infty (t-\mu)^2G(t)^2dt$. Then

$$I(\tilde{F})=\mathbb{P}\Big(\sum_{i=1}^k Z_i<1\Big),\qquad J_\ell(\tilde{F})\ge\Big(\int_0^{1/2}G(t)dt\Big)^2\mathbb{P}\Big(\sum_{i=1}^k Z_i<\frac{1}{2}\Big).$$

The random variable $\sum_{i=1}^k Z_i$ has mean $k\mu$ and variance $k\sigma^2$, and so it becomes concentrated on $k\mu$ when k is large provided $\sigma^2/\mu\to 0$ as $k\to\infty$. In particular, if $\mu<1/3k$ and $\sigma^2k\to 0$, then $\mathbb{P}(\sum_{i=1}^k Z_i<1/2)$ approaches 1 as $k\to\infty$. Therefore, to show that $\sum_{\ell=1}^k J_\ell(\tilde{F})/I(\tilde{F})\to\infty$ as $k\to\infty$, it suffices to find a function G satisfying

$$\int_0^\infty tG(t)^2<\frac{1}{3k},\qquad \int_0^\infty G(t)^2dt=1,$$

$$k\int_0^\infty t^2G(t)^2dt\to 0\quad\text{and}\quad k\Big(\int_0^\infty G(t)dt\Big)^2\to\infty\quad\text{as}\quad k\to\infty.$$

We find that choosing

$$G(t)\approx\begin{cases}\frac{\sqrt{k\log k}}{1+tk\log k}, & t<k^{-3/4},\\ 0, & \text{otherwise},\end{cases}\tag{3-3}$$

gives a function G satisfying these constraints. (This choice can be found via the calculus of variations, and further calculations show that this choice is essentially optimal.) Putting this all together, we find that for some constant $c>0$

$$\mathbb{E}\,\#\{i:\ L_i(n)\text{ prime}\}\ge(c\log k+o(1))\frac{\log R}{\log X}.$$

In particular, taking $R=X^{1/4-o(1)}$ (as allowed by the Bombieri-Vinogradov theorem) and letting k be sufficiently large, we find that there are infinitely many integers n such that $(c\log k)/4+o(1)$ of the $L_i(n)$ are simultaneously prime. Performing these calculations carefully allows one to take $c\approx 1$ when k is large.

Morally, the effect of such a choice of function $\tilde{F}$ is to make 'typical' divisors $(d_1, \dots, d_k)$ occurring in (3-1) to have $\prod_{i=1}^k d_i$ smaller, but for it to be more common that some of the components of $(d_1, \dots, d_k)$ are unusually large when compared with (2-1) or (2-3). This correspondingly causes the random integer n chosen with probability $w(n)$ to be such that the $L_i(n)$ are more likely to have slightly smaller prime factors, but it is also more likely that some of the $L_i(n)$ are prime.

4 Consequences for small gaps between primes

The above argument shows that there is a constant $c > 0$ such that for any set $\{L_1, \dots, L_k\}$ of integral linear functions satisfying the hypotheses of Conjecture 3, at least $c \log k$ of the linear functions L_i are simultaneously prime infinitely often. To show that there are primes close together, we simply take the linear functions to be of the form $L_i(n) = n + h_i$ for some integers $h_1 \le \cdots \le h_k$ chosen to make $\prod_{i=1}^k L_i(n)$ have no fixed prime divisor, and so that $h_k - h_1$ is small.

In general, a good choice of the h_i is to take h_i to be the i^{th} prime after the integer k. With this choice, $\prod_{i=1}^k L_i(n)$ has no fixed prime divisor and $h_k - h_i \approx k \log k$, so we can find intervals of length $k \log k$ containing $c \log k$ primes infinitely often. By working out the best possible implied constants, the argument we have sketched allows us to show that there are m primes in an interval of length $O(m^3 e^{4m})$ infinitely often. By incorporating refinements of the work of Zhang on primes in arithmetic progressions by the Polymath 8a project Castryck et al. [2014], and using some bounds from Harman's sieve Harman [2007], this can be improved slightly to $O(e^{3.815m})$ by work of R. C. Baker and Irving [2017].

If we are only interested in just how small a *single* gap can be, then we can improve the analysis and get an explicit bound on the size of the gap by adopting a numerical analysis perspective. The key issue is to find the smallest value of k such that for all large X

$$\mathbb{E}\,\#\{1 \le i \le k : L_i(n) \text{ prime}\} > 1,$$

since this immediately implies that two of the linear functions are simultaneously prime. Recalling that we can choose $R = x^{1/4-\epsilon}$ for any $\epsilon > 0$, using (3-2) we reduce the problem to finding a value of k as small as possible, such that we can find a function $\tilde{F} : [0, \infty)^k \to \mathbb{R}$ with

$$\frac{\sum_{i=1}^k J_i(\tilde{F})}{I(\tilde{F})} > 4.$$

We fix some basis functions $g_1, \dots, g_r : [0, \infty)^k \to \mathbb{R}$ which are supported on $(t_1, \dots, t_k)$ satisfying $\sum_{i=1}^k t_i \le 1$, and restrict our attention to functions $\tilde{F}$ in the linear span of the g_i; that is functions $\tilde{F}$ of the form

$$\tilde{F}(t_1,\dots,t_k)=\sum_{i=1}^{r} f_i g_i(t_1,\dots,t_r),$$

for some coefficients $\mathbf{f}=(f_1,\dots f_r)\in\mathbb{R}^r$ which we think of as variables we will optimize over. For such a choice, we find that $\sum_{i=1}^k J_i(\tilde{F})$ and $I(\tilde{F})$ are both quadratic forms in the variables $f_1,\dots,f_r$, and the coefficients of these quadratic forms are explicit integrals in terms of $g_1,\dots,g_r$. If we choose a suitably nice basis $g_1,\dots,g_r$, then these integrals are explicitly computable, and so we obtain explicit $r\times r$ real symmetric matrices M_1 and M_2 such that

$$\frac{\sum_{i=1}^k J_i(\tilde{F})}{I(\tilde{F})}=\frac{\mathbf{f}^T M_1\mathbf{f}}{\mathbf{f}^T M_2\mathbf{f}}.$$

We then find that the choice of coefficients $\mathbf{f}$ which maximizes this ratio is the eigenvector of $M_2^{-1}M_1$ corresponding to the largest eigenvalue, and the value of the ratio is given by this largest eigenvalue. Thus the existence of a good function $\tilde{F}$ can be reduced to checking whether the largest eigenvalue of a finite matrix is larger than 4, which can be performed numerically by a computer.

If we choose the g_i to be symmetric polynomials of low degree, then this provides a nice basis since the corresponding integrals have a closed form solution, and allows one to make arbitrarily accurate numerical approximations to the optimal function $\tilde{F}$ with enough computation. For the problem at hand, these numerical calculations are large but feasible. This approach ultimately allows one to show that if $k=54$ there is a function $\tilde{F}$ such that $\sum_{i=1}^{54} J_i(\tilde{F})/I(\tilde{F})>4$.

To turn this into small gaps between primes we need to choose the shifts h_i in our linear functions $L_i(n)=n+h_i$ so that $\prod_{i=1}^{54} L_i(n)$ has no fixed prime divisor, and the h_i are in as short an interval as possible. This is a feasible exhaustive numerical optimization problem, with an optimal choice of the $\{h_1,\dots,h_{54}\}$ given by [3]

$$\begin{aligned}\{&0,2,6,12,20,26,30,32,42,56,60,62,72,74,84,86,90,96,\\&104,110,114,116,120,126,132,134,140,144,152,156,\\&162,170,174,176,182,186,194,200,204,210,216,222,\\&224,230,236,240,242,246,252,254,260,264,266,270\}.\end{aligned}$$

Putting this together, we find

$$\liminf_n(p_{n+1}-p_n)\le 270.$$

[3] Such computations were first performed by Engelsma - see http://www.opertech.com/primes/k-tuples.html

This result is not quite the current record - in Polymath [2014] we make some further technical refinements to the sieve, which corresponds to modifying the expressions $J_i(\tilde{F})$ and $I(\tilde{F})$ slightly. This ultimately allows us to improve $k = 54$ to $k = 50$ and correspondingly improve from gaps of size at most 270 to gaps of at most 246, giving Theorem 2. The main ideas are the same as above.

5 Large gaps between primes

It turns out that because Theorem 1 gives strong partial information about the joint distribution of prime solutions to linear equations, which lie at the heart of many basic estimates about primes, it can be used to make progress on the existence of *large* gaps between primes, although the connection is less direct than with small gaps.

The first major breakthrough on large gaps between primes was due to Westzynthius [1931], who showed that there were gaps which could be arbitrarily large compared with the average gap. During the 1930s this was refined with ideas due to Erdős [1940] and Rankin [1936], giving

$$\sup_{p_n \le X} (p_{n+1} - p_n) \ge c' \frac{\log X \cdot \log\log X \cdot \log\log\log\log X}{(\log\log\log X)^2}, \tag{5-1}$$

for some constant $c' > 0$. Subsequent improvements over the next 75 years Schönhage [1963], Rankin [1962/1963], and Maier and Pomerance [1990] were only in improving the value of the constant c', the strongest being $c' = 2e^\gamma + o(1)$, due to Pintz [2014]. All of these approaches are based on a variant of the following lemma, which reduces the problem of constructing large gaps between primes to a combinatorial covering problem.

Lemma 5. *If one can choose residue classes $a_p \pmod p$ for $p \le x$ such that every element of $\{1, \dots, y\}$ is congruent to $a_p \pmod p$ for some $p \le x$, then there is a prime $p_n \ll e^x$ such that $p_{n+1} - p_n \ge y$.*

This lemma (which is a simple consequence of the Chinese Remainder Theorem), can be thought of as a natural generalization of the argument that $n! + 2, \dots, n! + n$ are n consecutive composite integers, and so explicitly demonstrates a prime gap of size at least n. (This roughly corresponds to choosing $a_p = 1$ for all primes p and taking $y = x$.)

The key idea in the Erdős–Rankin construction was to choose $a_p = 0$ for ‘medium sized’ primes, and choose a_p differently for small and large primes. Specifically, a version of their argument follows the following strategy to choose the a_p in turn, for some parameter $2 < z < x^{1/2}$:

1. Choose $a_p = 0$ for ‘medium primes’ $p \in [z, x/3]$.

2. Choose $a_p = 1$ for 'small primes' $p < z$.

3. Choose a_p greedily for 'large primes' $p \in (x/3, x]$.

By 'choosing greedily' we mean that we pick a residue class $a_q \pmod q$ which contains the largest number of integers in $\{1, \dots, y\}$ which are not congruent to $a_p \pmod p$ for some previously chosen a_p. There must be a residue class containing at least one uncovered element if we have not already covered all elements.

By choosing z appropriately[4], this allows one to cover $\{1, \dots, y\}$ by residue classes $a_p \pmod p$ for $p \le 7y(\log\log y)^2/(\log y \cdot \log\log\log y)$, which ultimately results in the bound (5-1). The key feature making this construction work is that choosing $a_p = 0$ for a very large number of 'medium primes' is much more efficient that a typical choice. This is because integers in $\{1, \dots, y\}$ with no prime factors bigger than z_1 are much less common than integers avoiding a random residue class for each prime bigger than z_1.

It was a well-known challenge of Erdős as to whether one could improve upon (5-1) by an arbitrarily large constant, and this was verified independently by Ford, Green, Konyagin, and Tao [2016] and the author Maynard [2016b]. Ultimately the approach of Ford-Green-Konyagin-Tao relied on the work of Green-Tao and Green-Tao-Ziegler on linear equations in primes Green and Tao [2010], Green and Tao [2012], and Green, Tao, and Ziegler [2012], whereas the work of the author relied on versions of Theorem 1. So far this second approach has proved more flexible for gaining quantitative improvements over (5-1), the strongest known results being due to a collaboration between all these authors Ford, Green, Konyagin, Maynard, and Tao [2018].

We focus here on the approach based around Theorem 1, first thinking about obtaining an arbitrarily large constant improvement over (5-1). We follow the same overall strategy as Erdős–Rankin, but improve the analysis for the large primes. By choosing the residue classes in a more sophisticated manner, we are able to remove *many* uncovered elements on average, rather than just 1 element, and this is the key feature which allows us to improve on (5-1).

Using the same choice of a_p as Erdős–Rankin for $p \le x/3$, we see that the elements of $\{1, \dots, y\}$ which are not covered by $a_p \pmod p$ for $p \le x/3$ are integers $n < y$ where n has no prime factors in $[z, x/3]$ and $n - 1$ has no prime factors less than z. Let us call the set of such integers $\mathcal{S}$. This is a set which is very similar to the primes, since most elements have a very large prime factor, and it is not clear that there are *any* possible a_p which might cover more than one additional element. Indeed, a typical residue class $a_p \pmod p$ for $p > x/3$ will contain *no* elements of $\mathcal{S}$. The problem of showing the existence of unusual residue classes containing many elements of $\mathcal{S}$ leads us to the following toy problem.

[4] $z = \exp(\log x \cdot \log\log\log x \,/\, 2\log\log x)$ gives a suitable choice.

Problem 1. *Given a prime modulus q, can we find a residue class $a_q \pmod q$ which contains many primes all less than $q(\log q)^{1/2}$?*

This toy problem can be answered in the positive by a variant of Theorem 1. If we choose our linear functions to be of the form $L_i(n) = n + h_i q$ for suitable constants h_i, then the existence of a residue class containing many small primes is implied by many of the L_i being simultaneously prime at n for some $n \le q$. The underlying sieve methodology is flexible enough to handle the fact that now the linear functions depend on q, and so can solve the toy problem. This correspondingly shows that for each large prime p, there are some residue classes $a_p \pmod p$ containing many elements of $\mathcal{S}$.[5]

To turn this into an actual covering, we need the residue classes for *different* large primes to be approximately 'independent' from one another. To do this we use the probabilistic method, by choosing a residue class at random for each large prime $p \in (x/3, x/2]$, and showing that with high probability this results in approximately independent behavior. Specifically, we choose a_p randomly with

$$\mathbb{P}(a_p = a \pmod p) \propto \sum_{\substack{n \equiv a \pmod p \\ L_1(n),\dots,L_k(n) \in \mathcal{S}}} w_{\mathcal{L}_p}(n),$$

where $w_{L_p}(n)$ is the normalized sieve weight introduced in Section 3 for the functions $L_1, \dots, L_k$ with $L_i(n) = n + h_i p$. We make these choices independently for all $p \in (x/3, x/2]$, so

$$\mathbb{P}\,(n \text{ not covered by large primes}) = \prod_{p \in [x/3, x/2]} \Big(1 - \sum_{m \equiv n \pmod p} w_{\mathcal{L}_p}(m)\Big)$$
$$\le \exp\Big(-\mathbb{E}\,\#\{p \in [x/3, x/2] : n \equiv a_p \pmod p\}\Big).$$

In particular, if the expected number of times any $n \in \mathcal{S}$ is congruent to $a_p \pmod p$ for some $p \in (x/3, x/2]$ is at least t, then the expected number of elements of $\mathcal{S}$ which are not covered by $a_p \pmod p$ for $p \le x/2$ is at most $e^{-t}\#\mathcal{S}$. If t is large, we can then greedily choose residue classes $a_p \pmod p$ for $p \in (x/2, x]$ to cover these few remaining elements.

Thus our new strategy for choosing the a_p is:

1. Choose $a_p = 0$ for 'medium primes' $p \in [z, x/3]$.

2. Choose $a_p = 1$ for 'small primes' $p < z$.

[5] The work of Green-Tao on linear equations in primes allows one to show for *most* large primes q there are k primes less than $q(\log q)^{1/2}$ for any fixed k, which is sufficient to improve the estimates for large primes.

3. Choose a_p randomly according to the sieve weights $w_{\mathcal{L}_p}$ independently for 'large primes' $p \in (x/3, x/2]$.

4. Choose a_p greedily for 'very large primes' $p \in (x/2, x]$.

This gives an arbitrarily large constant improvement over (5-1) provided we show that we can take t arbitrarily large. Calculations similar to those of Section 3 allow us to take t to be of size $\log k$ when we consider k linear functions, and so letting k be large enough we succeed in showing that the constant in (5-1) can indeed be taken to be arbitrarily large.

To get a quantitative improvement over (5-1) we run the essentially the same argument, but we need a version of Theorem 1 which has uniformity with respect to the number k of linear functions we consider as well as uniformity with respect to the coefficients of the linear functions. Such a version of Theorem 1 was established in Maynard [2016a]. With some technical modifications, this would ultimately yield a bound

$$\sup_{p_n \le X} (p_{n+1} - p_n) \ge c'' \frac{\log X \cdot \log\log X}{\log\log\log X}$$

for some constant $c'' > 0$. This is not quite as good as Theorem 3, because one can improve the quantitative argument further by being more careful about the manner in which we choose the a_p for different large primes. If instead of choosing residue classes independently at random we use ideas based on the 'semi-random' or 'Rödl nibble' method from combinatorics, we are able to establish a hypergraph covering lemma which allows our covering by residue classes to have almost no overlaps. After working through the technical details, this ultimately gives an additional improvement of a factor $\log\log\log\log X$, and hence gives Theorem 3.

6 Limitations

Both Theorem 1 and Theorem 3 appear to be the qualitative limit of what these methods can give, and require an entirely new approach to do better. Theorem 2 depends on the quantitative aspects of Theorem 1, and can potentially be improved slightly. New ideas are likely needed to significantly improve upon Theorem 2, however.

Optimal weights in high-dimensional sieves are poorly understood, and we do not know of general barriers beyond the parity phenomenon. In the context of Theorem 1 the parity phenomenon means that we cannot hope to prove $k/2$ of our linear functions are simultaneously prime based on a sieve argument. In particular, significant new ideas are required to attack the Twin Prime Conjecture.

Although in principle this leaves open the possibility of having rather more than $\log k$ of the linear functions being simultaneously prime, since the Selberg sieve weights seem to perform best in high dimensional sieving situations, we expect that it is unlikely to be possible to much do better than (3-2) based on weights formed by short divisor sums, even if we cannot prove a direct obstruction. Heuristic arguments show that over fairly general classes of potential Selberg sieve weights, we do not expect to do better than the choice given by (3-1). Given the choice (3-1), it is possible to show that a choice of smooth function similar to (3-3) is essentially best possible.

For the question of explicit small gaps between primes, there is some potential for further progress. Stronger results about primes in arithmetic progressions allow us to take R in (3-2) larger, which should reduce the critical value of k, and hence the size of the gap. The current record of 246 in Theorem 2 does not make use of the new equidistribution estimates of Zhang or its refinements (but these would likely only lead to small improvements). If we assume optimistic conjectures on the distribution of primes then we can do significantly better. Under the Elliott-Halberstam conjecture one can show gaps of size 12 Maynard [2015], and a generalization of the Elliott-Halberstam conjecture to numbers with several prime factors allows us to reach the absolute limit of these methods. Specifically, we have the following.

Theorem 6 (Polymath [2014]). *Assume the 'Generalized Elliott-Halberstam Conjecture' (see Polymath [ibid.]). Then we have*

$$\liminf_{n}(p_{n+1}-p_n)\leq 6.$$

The parity phenomenon makes it impossible for a first moment method of this type to prove a result less than 6, and so Theorem 6 is the strongest result of this type we can hope to prove along these lines.

All proofs showing the existence of large gaps between primes rely on some variant of Lemma 5, which allows one to construct a sequence of consecutive composite integers $n_1,\dots,n_1+y$ all with a prime factor of size $O(\log n_1)$. This places a severe limitation on how large the gaps we produce can be, since we expect that a large gap between primes will involve many composites n whose smallest prime factor is much larger than $\log n$. Specifically, Maier and Pomerance [1990] conjectured that the largest string of consecutive integers less than X all containing such a small prime factor should be of length

$$\log X\cdot(\log\log X)^{2+o(1)}.$$

Therefore we do not expect to be able to produce gaps larger than this without a new approach. The 'semi-random' method used in Ford, Green, Konyagin, Maynard, and Tao [2018] to show the existence of a good choice residue classes for large primes is essentially as good as one can hope for, so any quantitative progress based on the same overall method would likely require an improvement to Problem 1, showing the existence of more than $\log\log q$ primes less than $q(\log q)^{1/2}$ in some residue class modulo q. A uniform version of Conjecture 3 suggests that there should be residue classes containing roughly $(\log q)^{1/2}/\log\log q$ such primes, but we do not know how to prove this.

7 Other applications and further reading

For a more thorough survey of the details of these ideas on small gaps between primes, as well as the ideas behind the breakthrough of Zhang, we refer the reader to the excellent survey articles of Granville [2015] and Kowalski [2015]. For a more details of the original work of Goldston, Pintz and Yıldırım we refer the reader to survey by Soundararajan [2007].

One useful feature of the argument of Section 3 is that the full strength of the Bombieri-Vinogradov theorem was not required to prove bounded gaps between primes. Provided we can estimate primes in arithmetic progressions with modulus of size x^ϵ on average, we would obtain a version of Theorem 1 with $c\epsilon \log k$ of the linear functions are prime. This allows one to show the existence of bounded gaps between primes in many subsets of the primes where one has this type of weaker arithmetic information. A general statement of this type was established in Maynard [2016a]. The fact that one can restrict the entire argument to an arithmetic progression also allows one to get some control on the joint distribution of various arithmetic functions. There have been many recent works making use of these flexibilities in the setup of the sieve method, including Thorner [2014], Castillo, Hall, Lemke Oliver, Pollack, and Thompson [2015], Banks, Freiberg, and Turnage-Butterbaugh [2015], Freiberg [2016], Pollack [2014], H. Li and Pan [2015], R. C. Baker and Pollack [2016], Matomäki and Shao [2017], R. C. Baker and Zhao [2016a], Chua, Park, and Smith [2015], Vatwani [2017], Troupe [2016], Pintz [2015, 2017], Huang and Wu [2017], R. C. Baker and Zhao [2016b], Banks, Freiberg, and Maynard [2016], R. Baker and Freiberg [2016], Kaptan [2016], Parshall [2016], and Pollack and Thompson [2015].

New results on long gaps between primes have also found further applications to other situations R. Baker and Freiberg [2016], Maier and Rassias [2017], and J. Li, Pratt, and Shakan [2017]. It is hopeful that the ideas behind Theorem 1thrm:ManyGaps can find further applications in the future.

References

R. C. Baker and A. J. Irving (2017). "Bounded intervals containing many primes". *Math. Z.* 286.3-4, pp. 821–841. MR: 3671562 (cit. on p. 369).

Roger C. Baker and Paul Pollack (2016). "Bounded gaps between primes with a given primitive root, II". *Forum Math.* 28.4, pp. 675–687. MR: 3518382 (cit. on p. 376).

Roger C. Baker and Liangyi Zhao (2016a). "Gaps between primes in Beatty sequences". *Acta Arith.* 172.3, pp. 207–242. MR: 3460812 (cit. on p. 376).

– (2016b). "Gaps of smallest possible order between primes in an arithmetic progression". *Int. Math. Res. Not. IMRN* 23, pp. 7341–7368. MR: 3632085 (cit. on p. 376).

Roger Baker and Tristan Freiberg (2016). "Limit points and long gaps between primes". *Q. J. Math.* 67.2, pp. 233–260. MR: 3509992 (cit. on p. 376).

William D. Banks, Tristan Freiberg, and James Maynard (2016). "On limit points of the sequence of normalized prime gaps". *Proc. Lond. Math. Soc. (3)* 113.4, pp. 515–539. MR: 3556490 (cit. on p. 376).

William D. Banks, Tristan Freiberg, and Caroline L. Turnage-Butterbaugh (2015). "Consecutive primes in tuples". *Acta Arith.* 167.3, pp. 261–266. MR: 3316460 (cit. on p. 376).

E. Bombieri (1965). "On the large sieve". *Mathematika* 12, pp. 201–225. MR: 0197425 (cit. on p. 366).

E. Bombieri and H. Davenport (1966). "Small differences between prime numbers". *Proc. Roy. Soc. Ser. A* 293, pp. 1–18. MR: 0199165 (cit. on p. 366).

E. Bombieri, J. B. Friedlander, and H. Iwaniec (1986). "Primes in arithmetic progressions to large moduli". *Acta Math.* 156.3-4, pp. 203–251. MR: 834613 (cit. on p. 366).

– (1987). "Primes in arithmetic progressions to large moduli. II". *Math. Ann.* 277.3, pp. 361–393. MR: 891581 (cit. on p. 366).

– (1989). "Primes in arithmetic progressions to large moduli. III". *J. Amer. Math. Soc.* 2.2, pp. 215–224. MR: 976723 (cit. on p. 366).

Abel Castillo, Chris Hall, Robert J. Lemke Oliver, Paul Pollack, and Lola Thompson (2015). "Bounded gaps between primes in number fields and function fields". *Proc. Amer. Math. Soc.* 143.7, pp. 2841–2856. MR: 3336609 (cit. on p. 376).

Wouter Castryck et al. (2014). "New equidistribution estimates of Zhang type". *Algebra Number Theory* 8.9, pp. 2067–2199. MR: 3294387 (cit. on p. 369).

Lynn Chua, Soohyun Park, and Geoffrey D. Smith (2015). "Bounded gaps between primes in special sequences". *Proc. Amer. Math. Soc.* 143.11, pp. 4597–4611. MR: 3391020 (cit. on p. 376).

Harald Cramér (1936). "On the order of magnitude of the difference between consecutive prime numbers". *Acta Arithmetica* 2.1, pp. 23–46 (cit. on p. 363).

P. D. T. A. Elliott and H. Halberstam (1970). "A conjecture in prime number theory". In: *Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69)*. Academic Press, London, pp. 59–72. MR: 0276195 (cit. on p. 365).

P. Erdős (1940). "The difference of consecutive primes". *Duke Math. J.* 6, pp. 438–441. MR: 0001759 (cit. on pp. 366, 371).

Kevin Ford, Ben Green, Sergei Konyagin, James Maynard, and Terence Tao (2018). "Long gaps between primes". *J. Amer. Math. Soc.* 31.1, pp. 65–105. MR: 3718451 (cit. on pp. 364, 372, 375).

Kevin Ford, Ben Green, Sergei Konyagin, and Terence Tao (2016). "Large gaps between consecutive prime numbers". *Ann. of Math. (2)* 183.3, pp. 935–974. MR: 3488740 (cit. on p. 372).

E. Fouvry and H. Iwaniec (1980). "On a theorem of Bombieri-Vinogradov type". *Mathematika* 27.2, 135–152 (1981). MR: 610700 (cit. on p. 366).

Tristan Freiberg (2016). "Short intervals with a given number of primes". *J. Number Theory* 163, pp. 159–171. MR: 3459566 (cit. on p. 376).

Daniel A. Goldston, János Pintz, and Cem Y. Yıldırım (2009). "Primes in tuples. I". *Ann. of Math. (2)* 170.2, pp. 819–862. MR: 2552109 (cit. on pp. 364, 366).

Andrew Granville (2015). "Primes in intervals of bounded length". *Bull. Amer. Math. Soc. (N.S.)* 52.2, pp. 171–222. MR: 3312631 (cit. on p. 376).

Ben Green and Terence Tao (2012). "The Möbius function is strongly orthogonal to nilsequences". *Ann. of Math. (2)* 175.2, pp. 541–566. MR: 2877066 (cit. on p. 372).

Ben Green, Terence Tao, and Tamar Ziegler (2012). "An inverse theorem for the Gowers $U^{s+1}[N]$-norm". *Ann. of Math. (2)* 176.2, pp. 1231–1372. MR: 2950773 (cit. on p. 372).

Benjamin Green and Terence Tao (2010). "Linear equations in primes". *Ann. of Math. (2)* 171.3, pp. 1753–1850. MR: 2680398 (cit. on p. 372).

Glyn Harman (2007). *Prime-detecting sieves*. Vol. 33. London Mathematical Society Monographs Series. Princeton University Press, Princeton, NJ, pp. xvi+362. MR: 2331072 (cit. on p. 369).

Wen Huang and XiaoSheng Wu (2017). "On the set of the difference of primes". *Proc. Amer. Math. Soc.* 145.9, pp. 3787–3793. MR: 3665033 (cit. on p. 376).

M. N. Huxley (1973). "Small differences between consecutive primes". *Mathematika* 20, pp. 229–232. MR: 0352021 (cit. on p. 366).

– (1977). "Small differences between consecutive primes. II". *Mathematika* 24.2, pp. 142–152. MR: 0466042 (cit. on p. 366).

Martin Huxley (1984). "An application of the Fouvry-Iwaniec theorem". *Acta Arith.* 43.4, pp. 441–443. MR: 756293 (cit. on p. 366).

Deniz A. Kaptan (2016). "A note on small gaps between primes in arithmetic progressions". *Acta Arith.* 172.4, pp. 351–375. MR: 3471048 (cit. on p. 376).

Emmanuel Kowalski (2015). "Gaps between prime numbers and primes in arithmetic progressions [after Y. Zhang and J. Maynard]". *Astérisque* 367-368, Exp. No. 1084, ix, 327–366. MR: 3363595 (cit. on p. 376).

Hongze Li and Hao Pan (2015). "Bounded gaps between primes of a special form". *Int. Math. Res. Not. IMRN* 23, pp. 12345–12365. MR: 3431623 (cit. on p. 376).

Junxian Li, Kyle Pratt, and George Shakan (2017). "A lower bound for the least prime in an arithmetic progression". *Q. J. Math.* 68.3, pp. 729–758. MR: 3698292 (cit. on p. 376).

Helmut Maier (1988). "Small differences between prime numbers". *Michigan Math. J.* 35.3, pp. 323–344. MR: 978303 (cit. on p. 366).

Helmut Maier and Carl Pomerance (1990). "Unusually large gaps between consecutive primes". *Trans. Amer. Math. Soc.* 322.1, pp. 201–237. MR: 972703 (cit. on pp. 371, 375).

Helmut Maier and Michael Th. Rassias (2017). "Large gaps between consecutive prime numbers containing perfect k-th powers of prime numbers". *J. Funct. Anal.* 272.6, pp. 2659–2696. MR: 3603310 (cit. on p. 376).

Kaisa Matomäki and Xuancheng Shao (2017). "Vinogradov's three primes theorem with almost twin primes". *Compos. Math.* 153.6, pp. 1220–1256. MR: 3647436 (cit. on p. 376).

James Maynard (2015). "Small gaps between primes". *Ann. of Math. (2)* 181.1, pp. 383–413. MR: 3272929 (cit. on pp. 364, 367, 375).

– (2016a). "Dense clusters of primes in subsets". *Compos. Math.* 152.7, pp. 1517–1554. MR: 3530450 (cit. on pp. 364, 374, 376).

– (2016b). "Large gaps between primes". *Ann. of Math. (2)* 183.3, pp. 915–933. MR: 3488739 (cit. on p. 372).

Hans Parshall (2016). "Small gaps between configurations of prime polynomials". *J. Number Theory* 162, pp. 35–53. MR: 3448260 (cit. on p. 376).

JZ Pilt'ja (1972). "The magnitude of the difference between consecutive primes". *Studies in number theory*. Izdat. Saratov. Univ., Saratov (Russian), pp. 73–79 (cit. on p. 366).

János Pintz (2014). "Some new results on gaps between consecutive primes". In: *Number theory, analysis, and combinatorics*. De Gruyter Proc. Math. De Gruyter, Berlin, pp. 261–278. MR: 3220124 (cit. on p. 371).

– (2015). "On the ratio of consecutive gaps between primes". In: *Analytic number theory*. Springer, Cham, pp. 285–304. MR: 3467403 (cit. on p. 376).

– (2017). "Patterns of primes in arithmetic progressions". In: *Number theory – Diophantine problems, uniform distribution and applications*. Springer, Cham, pp. 369–379. MR: 3676411 (cit. on p. 376).

Paul Pollack (2014). "Bounded gaps between primes with a given primitive root". *Algebra Number Theory* 8.7, pp. 1769–1786. MR: 3272281 (cit. on p. 376).

Paul Pollack and Lola Thompson (2015). "Arithmetic functions at consecutive shifted primes". *Int. J. Number Theory* 11.5, pp. 1477–1498. MR: 3376222 (cit. on p. 376).

D. H. J. Polymath (2014). "Variants of the Selberg sieve, and bounded intervals containing many primes". *Res. Math. Sci.* 1, Art. 12, 83. MR: 3373710 (cit. on pp. 364, 371, 375).

R. A. Rankin (1936). "The Difference between Consecutive Prime Numbers". *J. London Math. Soc.* 11.4, pp. 242–245. MR: 1574971 (cit. on p. 371).

– (1950). "The difference between consecutive prime numbers. IV". *Proc. Amer. Math. Soc.* 1, pp. 143–150. MR: 0034800 (cit. on p. 366).

– (1962/1963). "The difference between consecutive prime numbers. V". *Proc. Edinburgh Math. Soc. (2)* 13, pp. 331–332. MR: 0160767 (cit. on p. 371).

Giovanni Ricci (1954). "Sull'andamento della differenza di numeri primi consecutivi". *Riv. Mat. Univ. Parma* 5, pp. 3–54. MR: 0067139 (cit. on p. 366).

Arnold Schönhage (1963). "Eine Bemerkung zur Konstruktion grosser Primzahllücken". *Arch. Math. (Basel)* 14, pp. 29–30. MR: 0146154 (cit. on p. 371).

K. Soundararajan (2007). "Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım". *Bull. Amer. Math. Soc. (N.S.)* 44.1, pp. 1–18. MR: 2265008 (cit. on p. 376).

Jesse Thorner (2014). "Bounded gaps between primes in Chebotarev sets". *Res. Math. Sci.* 1, Art. 4, 16. MR: 3344173 (cit. on p. 376).

Lee Troupe (2016). "Bounded gaps between prime polynomials with a given primitive root". *Finite Fields Appl.* 37, pp. 295–310. MR: 3426591 (cit. on p. 376).

S. Uchiyama (1975). "On the difference between consecutive prime numbers". *Acta Arith.* 27. Collection of articles in memory of JuriĭVladimirovič Linnik, pp. 153–157. MR: 0366839 (cit. on p. 366).

Akshaa Vatwani (2017). "Bounded gaps between Gaussian primes". *J. Number Theory* 171, pp. 449–473. MR: 3556693 (cit. on p. 376).

A. I. Vinogradov (1965). "The density hypothesis for Dirichet L-series". *Izv. Akad. Nauk SSSR Ser. Mat.* 29, pp. 903–934. MR: 0197414 (cit. on p. 366).

Erik Johan Westzynthius (1931). "Über die Verteilung der Zahlen: die zu den n ersten Primzahlen teilerfremd sind." *Commentat. Phys.-Math.* 5.25, pp. 1–37 (cit. on p. 371).

Yitang Zhang (2014). "Bounded gaps between primes". *Ann. of Math. (2)* 179.3, pp. 1121–1174. MR: 3171761 (cit. on p. 366).

Received 2017-11-27.

James Maynard
Mathematical Institute
Oxford
England OX1 4AU
james.alexander.maynard@gmail.com

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (381–394)

THE SUBCONVEXITY PROBLEM FOR L-FUNCTIONS

Ritabrata Munshi

Abstract

Estimating the size of automorphic L-functions on the critical line is a central problem in analytic number theory. An easy consequence of the standard analytic properties of the L-function is the convexity bound, whereas the generalised Riemann Hypothesis predicts a much sharper bound. Breaking the convexity barrier is a hard problem. The moment method has been used to surpass convexity in the case of L-functions of degree one and two. In this talk I will discuss a different method, which has been quite successful to settle certain longstanding open problems in the case of degree three.

At the 1994 International Congress at Zürich, J. B. Friedlander [1995] briefly described the essence of the amplified moment method which he was developing in a series of joint works with Duke and Iwaniec, with the aim of obtaining non-trivial bounds for L-functions. Since then the amplification technique has proved to be very effective in a number of scenarios involving $GL(2)$ L-functions (see J. Friedlander and Iwaniec [1992], Duke, J. B. Friedlander, and Iwaniec [1993, 1994, 1995, 2001, 2002], Kowalski, Michel, and VanderKam [2002], Michel [2004], Harcos and Michel [2006], and Blomer and Harcos [2008]). But there are major hurdles in extending the method far beyond. In the last decade the automorphic period approach has been developed in great detail and generality (over number fields), by Michel, Venkatesh and others (see Bernstein and Reznikov [2010], Michel and Venkatesh [2010], Wu [2014]). This puts the moment method in a proper perspective and gives a satisfactory explanation to the 'mysterious identities between families of L-functions' that already occurs in the study of the moments of the Rankin-Selberg L-functions Harcos and Michel [2006], Michel [2004]. This has been the topic of Michel's address at the 2006 International Congress at Madrid Michel and Venkatesh [2006]. Here I will briefly describe a new approach to tackle subconvexity, which has not only settled some of the longstanding open problems in the field, but has also matched in strength the existing benchmarks. As there are several excellent accounts

MSC2010: primary 11F66; secondary 11M41.
Keywords: L-functions, subconvexity, circle method.

on the subconvexity problem for general automorphic L-functions and on its importance in number theory, equidistribution and beyond (see J. B. Friedlander [1995], Iwaniec and Sarnak [2000], Michel [2007], Michel and Venkatesh [2006], Sarnak [1998]), I will discuss some specific cases which will easily bring out the new features of the method in contrast to the amplified moment method of Friedlander-Iwaniec.

First let us recall the subconvexity problem and the amplification technique with the aid of an example. The Ramanujan Δ-function

$$\Delta(z) = \eta(z)^{24} = \sum_{n=1}^{\infty} \tau(n) e^{2\pi i n z}$$

is a modular form of weight 12 for the full modular group $SL(2,\mathbb{Z})$, and is the prototype for all modular forms. Let χ be a primitive Dirichlet character with modulus M, then the twist $\Delta \otimes \chi$ is a modular form of weight 12 for the congruence group $\Gamma_0(M^2)$ with nebentypus χ^2. It was conjectured by Ramanujan, and later proved by Deligne, that $\tau(n) \ll n^{\frac{11}{2}+\varepsilon}$. Accordingly we define the normalized Fourier coefficients $\lambda_\Delta(n) = \tau(n)/n^{11/2}$, so that the associated Hecke L-function

$$L(s, \Delta \otimes \chi) = \sum_{n=1}^{\infty} \frac{\lambda_\Delta(n)\chi(n)}{n^s}$$

is absolutely convergent for $\mathrm{Re}(s) > 1$ and satisfies the Riemann type functional equation $s \to 1-s$ with center at $s = 1/2$. The multiplicativity of the τ function, again conjectured by Ramanujan and proved shortly thereafter by Mordell (and by Hecke in general), leads to a degree two Euler product representation of this L-function. The basic analytic properties of this type of L-functions - analytic continuation, functional equation - were established by Hecke. This is an example of an automorphic L-function of degree two. In general, understanding the behaviour of automorphic L-functions inside the critical strip $0 \leq \mathrm{Re}(s) \leq 1$, and in particular on the central line $\mathrm{Re}(s) = 1/2$, is the main problem in this field.

There are few results in complex analysis which dictate the behaviour of holomorphic functions inside a domain, once their behaviour is known on the boundary. One such result is the convexity principle of Phragmén-Lindelöf, which when applied to the above L-function, yields the bound

$$L\left(\tfrac{1}{2} + it, \Delta \otimes \chi\right) \ll C^{1/4+\varepsilon}$$

for any $t \in \mathbb{R}$ and any $\varepsilon > 0$, where $C = [M(1+|t|)]^2$. (The notation '$\ll$' here means that there exists a constant $c(\varepsilon)$ depending only on ε such that the absolute value of the left-hand side is smaller than $c(\varepsilon)$ times the right-hand side.) The same bound can be obtained through the approximate functional equation which gives a Dirichlet series approximation to the L-value. It is now understood that the length of this approximation can not be smaller than the square-root of the size of the analytic conductor J. B. Friedlander [1995], Iwaniec and Sarnak [2000], which for the particular example we are considering is given by C. A direct consequence of the approximate functional equation is the bound

$$L(\tfrac{1}{2}+it, \Delta \otimes \chi) \ll C^{\varepsilon} \sup_{N \ll C^{1/2+\varepsilon}} \frac{|S(N)|}{N^{1/2}} + C^{-2018},$$

where $S(N)$ are Dirichlet polynomials of the form

$$S(N) = \sum_{n=1}^{\infty} \lambda_{\Delta}(n)\chi(n)n^{it}\, W\left(\frac{n}{N}\right)$$

with W a smooth bump function. A trivial estimation of this sum recovers the above convexity bound $O(C^{1/4+\varepsilon})$. This is far from what one expects to be the truth. Indeed the Generalized Riemann Hypothesis (GRH) implies the Generalized Lindelöf Hypothesis which predicts a bound with exponent 0 in place of $1/4$. Any bound with exponent $1/4 - \delta$ for some $\delta > 0$ is called a subconvex bound. Such bounds have several striking applications Michel [2007], Sarnak [1995].

In the example we are considering there are two distinct parameters of interest, namely t, which is allowed to take any real value, and M which can take any positive integral value. In other words, there are two types of subfamilies of L-values of interest, viz.

$$\{L(\tfrac{1}{2}+it, \Delta \otimes \chi) : t \in \mathbb{R}\}$$

where the character χ is kept fixed, and

$$\{L(\tfrac{1}{2}+it, \Delta \otimes \chi) : \chi \bmod M \;\; \text{primitive}, M \in \mathbb{N}\}$$

where t is held fixed and the character varies with the modulus tending to infinity. A subconvex bound for the former type of subfamilies is called t-aspect subconvexity and that for the latter type is called M-aspect (or twist aspect). Often, especially in arithmetic applications, the parameter M is of interest, and one is required to the break the convexity barrier in the M-aspect only. One such application is the uniform distribution of rational points on sphere (Linnik's problem) Michel [2007].

The t-aspect subconvexity for $L(s,\Delta)$ (χ principal character) was first established by Good using the spectral theory of Maass forms. Good's approach is via the moment method and is based on getting a strong error term in the asymptotic expansion for the second moment

$$\int_T^{2T} |L(1/2+it,\Delta)|^2 \mathrm{d}t.$$

In Good [1982] he establishes that the above integral is asymptotically

$$TP(\log T) + O((T\log T)^{2/3})$$

where P is a linear polynomial. This strong error term is achieved by studying the more concentrated second moment

$$\int_{T-T/U}^{T+T/U} |L(1/2+it,\Delta)|^2 \mathrm{d}t.$$

where one strives to get U as big as possible. Expanding the absolute value square using the approximate functional equation, and then executing the t-integral one is left with a shifted convolution sum problem which may be tackled by studying the Dirichlet series

$$\sum_{n=1}^{\infty} \frac{\tau(n)\overline{\tau(n+h)}}{(n+h/2)^s}.$$

The automorphic forms enter the picture as the above series can be realised as the Petersson inner product of the form Δ and the h-th Poincare series. The analytic continuation of this L-series beyond the region of absolute convergence was first obtained by Selberg [1965]. But it was Good who first effectively used the spectral interpretation of the above Dirichlet series to obtain estimates for moments of L-function. The estimates obtained by Good give the optimal choice $U=(T\log T)^{1/3}$, and the above asymptotic expansion follows. This in turn yields

$$L(\tfrac{1}{2}+it,\Delta) \ll (1+|t|)^{1/3+\varepsilon},$$

which is the $GL(2)$ analogue of the famous estimate of Hardy-Littlewood-Weyl for the Riemann zeta function $\zeta(1/2+it) \ll (1+|t|)^{1/6+\varepsilon}$. The most straightforward way to prove the Weyl bound for the zeta function is through exponential sums, but a proof based on the moment method, similar in spirit as above, was obtained by Iwaniec [1980].

The twist aspect subconvexity for $L(s,\Delta\otimes\chi)$ is a harder problem. One reason being that there is no simple way to recover a subconvex bound for an individual L-value from an asymptotic for the second moment

$$\sum_{\psi \bmod M} |L(1/2+it, \Delta \otimes \psi)|^2.$$

There is no natural way (while retaining some sort of spectral completeness of the family) to shorten the outer sum so as to obtain a second moment concentrated around χ, the particular character we are interested in. (However this is again possible for the weight/spectral aspect, e.g. Lau, Liu, and Ye [2006], Sarnak [2001].) Of course, one way will be to estimate a higher moment, say the fourth, but this turns out to be a much harder problem. Though one would not need an asymptotic expansion in this case, but even getting an appropriately strong bound is difficult. The amplification technique originated to bypass this problem. The simple idea being to consider the weighted moment

$$\sum_{\psi \bmod M} w(\psi)|L(1/2+it, \Delta \otimes \psi)|^2,$$

where the weights are necessarily non-negative, and $w(\chi)$ is in some sense larger than the average weight. One way to assign such weights (amplifier) is to consider sums of the form

$$w(\psi) = \left|\sum_{\ell \sim L} a(\ell)\psi(\ell)\right|^2,$$

where the coefficients $a(\ell)$ are allowed to depend on χ, Δ and t. In the particular example we are looking at, the choice is rather simple $a(\ell) = \bar{\chi}(\ell)$. Indeed $|\chi(\ell)| = 1$, hence bounded away from 0, if $(\ell, M) = 1$. (But such simple effective lower bounds are not available for Fourier coefficients of modular or Maass forms, and the construction of the amplifier, in the more general setup, has to go through deeper arithmetic structure like Hecke relations.) Using this amplifier Duke, J. B. Friedlander, and Iwaniec [1993] obtained the subconvex bound

$$L(1/2, \Delta \otimes \chi) \ll M^{\frac{1}{2}-\frac{1}{22}+\varepsilon}.$$

This bound has been improved and extended to cover twists of any $GL(2)$ automorphic form. The strongest bound is due to Blomer and Harcos [2008], where they get the exponent $1/2 - 1/8$ pushing to the limit the amplification method and utilising ideas of Bykovskii. This exponent corresponds to the classic result of Burgess $L(1/2, \chi) \ll M^{3/16+\varepsilon}$ for the Dirichlet L-function Burgess [1963]. In the same paper Blomer and Harcos get a hybrid subconvex bound

$$L(\tfrac{1}{2}+it, \Delta \otimes \chi) \ll [M(1+|t|)]^{\frac{1}{2}-\frac{1}{40}+\varepsilon}.$$

A strong hybrid subconvex bound for Dirichlet L-function was obtained by Heath-Brown [1980] extending the work of Burgess [1963].

In general, the basic philosophy of the moment method and the amplification technique can be described as follows. Suppose one seeks to get a bound for $L(\pi_0) = L(1/2, \pi_0)$ where π_0 is an automorphic form. The approximate functional equation reduces the problem to getting cancellation in a sum of the form

$$S_{\pi_0} = \sum_{n \sim C^{1/2}} \lambda_{\pi_0}(n)$$

where λ_{π_0} are the Whittaker-Fourier coefficients and C is the conductor of π_0. As a first step, we need to find a 'spectrally complete family' $\mathfrak{F}$ containing π_0, where all the objects in $\mathfrak{F}$ have 'comparable conductors'. Let

$$\mathfrak{M} = \sum_{\pi \in \mathfrak{F}} |S_\pi|^2,$$

then dropping all terms in the sum except the particular object we are interested in, we conclude the bound $S_{\pi_0} \ll \mathfrak{M}^{1/2}$. Since $\mathfrak{M}$ cannot be smaller than $|\mathfrak{F}|$, the diagonal contribution, for subconvexity we at least need that $|\mathfrak{F}| \ll C^{1/2-\delta}$. So the family cannot be too big. Indeed, larger the family less likely it is, even with amplification, to obtain a non-trivial bound for a particular L-value $L(\pi_0)$, as the individual contribution gets washed out in the large average. Now let us consider the problem of estimating the moment $\mathfrak{M}$. The usual way for the moment method is to open the absolute value and then execute the sum over forms

$$\sum_{\pi \in \mathfrak{F}} \lambda_\pi(n)\overline{\lambda_\pi(m)},$$

using some sort of quasi-orthogonality (e.g. Petersson trace formula). So to get a good estimate we need the family to be big enough - e.g. the off-diagonal contribution in the Petersson trace formula is easily seen to be smaller for larger families, say when the level is bigger. This dichotomy on the size of the family puts a severe restriction on the choice of $\mathfrak{F}$. It is probably 'an accident' that such families exist for L-functions of degree one and two. The amplification technique kicks in when one stumbles upon a family, where $|\mathfrak{F}| = C^{1/2}$. Then one has to manufacture a suitable amplifier A_π, and has to estimate the amplified moment

$$\mathfrak{M}^\sharp = \sum_{\pi \in \mathfrak{F}} |A_\pi|^2 |S_\pi|^2,$$

instead of $\mathfrak{M}$.

I shall now explain a different approach to subconvexity in the context of the above example. We are seeking cancellation in the sum $S(N) = \sum_{n\sim N} \lambda_\Delta(n)\chi(n)n^{it}$, where $N = C^{1/2}$. The basic idea is to separate the oscillation of the Fourier coefficient $\lambda_\Delta(n)$ from that of $\chi(n)n^{it}$. We do this bluntly by introducing a new variable, and rewriting the sum as

$$S(N) = \sum\sum_{\substack{m,n\sim N\\ m=n}} \lambda_\Delta(n)\chi(m)m^{it}.$$

Next we detect the equation $m = n$ using the circle method to arrive at

$$S(N) = \frac{1}{Q^2}\sum_{q\sim Q}\ \sum_{\substack{a \bmod q\\ (a,q)=1}} \Big[\sum_{n\sim N} \lambda_\Delta(n)e_q(an)\Big]\Big[\sum_{m\sim N} \chi(m)m^{it}e_q(-am)\Big]$$

with $Q = N^{1/2} = C^{1/4}$, where we are using the standard shorthand notation $e_q(z) = e^{2\pi i z/q}$. Summation formulas of Poisson and Voronoi, reduces the expression to a bilinear form with Kloosterman fractions

$$\sum\sum_{\substack{m,q\sim Q\\ (m,q)=1}} \chi(q\bar{m})(qm^{-1})^{it}e_q(-M\bar{m})$$

where $\bar{m}$ is the multiplicative inverse of m modulo q. The trivial estimation of this sum yields the convexity bound. A deep result of Duke, Friedlander and Iwaniec gives some cancellation in such sums. But in Munshi [2014] I proceeded in a different direction which paved the way for further developments in the method. Suppose by some means we had reached a similar expression with a factorizable modulus $q = q_1q_2$ with $q_i \sim Q_i$ and $Q_1Q_2 = Q$. Then one could apply the Cauchy inequality to dominate the above expression by

$$Q^{1/2}\sum_{q_2\sim Q_2}\left[\sum_{\substack{m\sim Q\\ (m,q_2)=1}}\Big|\sum_{\substack{q_1\sim Q_1\\ (m,q_1)=1}} \chi(q_1)q_1^{it}e_{q_1q_2}(-M\bar{m})\Big|^2\right]^{1/2},$$

and now it would be possible to get more cancellation by opening the absolute value square and applying the Poisson summation formula on the sum over m. The new input that one would need is the Weil bound for Kloosterman sums. But how does one get this desired structure on the modulus? In Munshi [ibid.] I used Jutila's version of the circle method to achieve this goal, and was able to prove

$$L(\tfrac{1}{2}+it, \Delta\otimes\chi) \ll [M(1+|t|)]^{\frac{1}{2}-\frac{1}{18}+\varepsilon},$$

which was an improvement over the above mentioned result of Blomer-Harcos. There is another interesting way to achieve the same structure - 'congruence-equation trick'. Pick a set of primes $\mathbb{Q}_1$ from the interval $[Q_1, 2Q_1]$. Then for any $q_1 \in \mathbb{Q}_1$ we split the integral equation $m = n$ into a congruence $m \equiv n \bmod q_1$ and a smaller integral equation $(m-n)/q_1 = 0$. The last equation can be detected by a circle method with modulus $Q_2 = \sqrt{N/Q_1}$. Detecting the congruence using additive characters modulo q_1, we arrive at an expression similar to one we got above with $q = q_1 q_2$. Of course the price we pay to get this factorization is the increase in the size of the modulus $q \sim \sqrt{Q_1 N}$. But the structural advantage compensates this loss adequately, and even provides the desired extra saving. (For applications of this trick to Diophantine problems, see Browning and Munshi [2013] and Munshi [2015a].)

Unfortunately, this simple approach does not work for the twisted $GL(3)$ L-functions $L(s, \pi \otimes \chi)$ where π is a Hecke-Maass cusp form for $SL(3, \mathbb{Z})$. Previously Li [2011] and Blomer [2012] had studied the t-aspect and the M-aspect subconvexity problems for these L-functions in the special case where π is a symmetric square lift of a $SL(2, \mathbb{Z})$ Hecke-Maass cusp form. (A p-adic version of Li's result was established in Munshi [2013b] using the ideas of Munshi [2013a].) Their approach is an extension of Conrey and Iwaniec [2000] method where non-negativity of certain L-values plays a crucial role. As such their results could not be extended to cover generic $SL(3, \mathbb{Z})$ forms. In Munshi [2015d] I partially succeeded in the M-aspect. Suppose the character χ factorises as $\chi_1 \chi_2$ where χ_i is primitive modulo M_i with $(M_1, M_2) = 1$, $M_1 < M_2$. We are seeking cancellation in the sum

$$S_3(N) = \sum_{n \sim N} \lambda_\pi(n, 1) \chi(n),$$

where $\lambda_\pi(n, r)$ are the normalised Whittaker-Fourier coefficients of the form π, and $N = M^{3/2}$ - square-root of the size of the conductor. As before we separate the oscillation of the Fourier coefficients from that of the character by introducing a new variable and an equation,

$$S_3(N) = \mathop{\sum\sum}_{\substack{m,n \sim N \\ m=n}} \lambda_\pi(n, 1) \chi(m).$$

Now we use the congruence-equation trick to split the integral equation $m = n$ as $m \equiv n \bmod M_1$ and the integral equation $(m-n)/M_1 = 0$. Here this trick acts as a level lowering mechanism as the modulus M_1 was intrinsic to the problem. The remaining integral equation is detected using the circle method with modulus of smaller size $Q =$

$\sqrt{N/M_1}$. The resulting expression now looks like

$$S_3(N) = \frac{1}{Q^2 M_1} \sum_{q \sim Q} \sum_{\substack{a \bmod qM_1 \\ (a,q)=1}} \Big[\sum_{n \sim N} \lambda_\pi(n,1) e_{qM_1}(an) \Big] \times \Big[\sum_{m \sim N} \chi(m) e_{qM_1}(-am) \Big],$$

and one is again able to win, as long as $\sqrt{M_2} < M_1 < M_2$, by applying summation formulas - Poisson summation and $GL(3)$ Voronoi summation - followed by an application of Cauchy to escape from the trap of involution, and another application of Poisson summation. Here one needs to use Deligne's bound for complete exponential sums.

Though it was clear that this approach would not extend to general characters, the t-aspect subconvexity for $L(s,\pi)$ for π a $SL(3,\mathbb{Z})$ form, became tractable. Indeed the t-aspect is related to twists by highly factorizable characters. In Munshi [2015b] I established the following subconvex bound.

Theorem 1. *Let π be a Hecke-Maass cusp form for $SL(3,\mathbb{Z})$. Then we have*

$$L(\tfrac{1}{2}+it,\pi) \ll (1+|t|)^{\frac{3}{4}-\frac{1}{16}+\varepsilon}.$$

Curiously the exponent matches with Li's bound in Li [2011] for symmetric square lifts, though the two approaches are totally different. The above result is proved using the technique outlined above, but now one have to use the archimedean analogue of the congruence-equation trick. Imagine $M_1 = p^r$ with p a fixed prime and $r \to \infty$, then the congruence condition $m \equiv n \bmod M_1$ corresponds to a condition on the p-adic size of $m-n$. This translates as $|m-n| \ll N/M_1$ in the archimedean situation of t-aspect. In Munshi [2015b] we choose a suitable parameter V and factorise the integral equation $m-n=0$ (of size N) into a distance condition $|m-n| \ll N/V$ and the smaller integral equation $m-n=0$ (of size N/V). The size restriction $|m-n| \ll N/V$ is then detected using an integral involving $(m/n)^{iv}$.

For the twist aspect one needs to introduce higher order harmonics. The usual circle methods and the DFI delta method are based on $GL(1)$ harmonics, or the harmonics of the abelian circle group S^1. These are the trigonometric functions $e(z)$. The delta method gives a Fourier resolution of the delta symbol $\delta : \mathbb{Z} \to \{0,1\}$, $\delta(0)=1$ and $\delta(n)=0$ for $n \neq 0$. A very rough version of this formula looks like

$$\delta(n) \approx \frac{1}{C^2} \sum_{c \sim C} \sum_{\substack{a \bmod c \\ (a,c)=1}} e_c(an).$$

Usually to detect the event $n = 0$ with n varying in the range $[-N, N]$ one takes $C = \sqrt{N}$ so as to minimise the total conductor - the arithmetic modulus c and the amplitude of the oscillation in the weight function (which we have ignored above) - in the circle method formula. In several cases, it turns out that all one needs is a circle method formula with C slightly smaller than $\sqrt{N}$. For example, the subconvexity of $L(s, \Delta \otimes \chi)$ follows quite easily if one has such an elusive circle method.

In the above applications we always needed a souped up version of the usual circle method formula, which we achieved by putting extra structural conditions on the fractions a/c that parametrise the outer sum. In all cases we ended up relating the L-value under focus to an average of products of L-values, e.g.

$$L(1/2 + it, \Delta \otimes \chi) \leftrightsquigarrow \sum_{f \in \mathfrak{F}} L(s_1, \Delta \otimes v_f) L(s_2 + it, \chi \otimes v_f)$$

where $\mathfrak{F}$ is a collection of (Farey) fractions with suitable factorization of denominator, and for $f = (a, c)$, v_f is the additive character $n \to e_c(an)$. Hence at the end we are still computing moments of certain products of L-values (but not sizes of L-values). Since clearly the $GL(1)$ harmonics will not suffice for higher degree L-functions, one looks for expansions of delta involving higher order harmonics. The trace formulas of non-abelian groups are natural sources for such expansions. For example, the Petersson trace formula for modular forms gives

$$\begin{aligned}\delta(m-n) = &\sum_{f \in H_k(q,\psi)} w_f^{-1} \lambda_f(n) \overline{\lambda_f(m)} \\ &- 2\pi i^{-k} \sum_{c=1}^{\infty} \frac{S_\psi(m,n;cq)}{cq} J_{k-1}\left(\frac{4\pi\sqrt{nm}}{cq}\right)\end{aligned}$$

for $m, n > 0$. Here $H_k(q, \psi)$ is an orthogonal Hecke basis for the space of cusp forms of weight k level q and nebentypus ψ, $S_\psi(m, n; c)$ are Kloosterman sums and $J_{k-1}(x)$ is the J-Bessel function. Since the left hand side does not depend on ψ, q or k, one may take suitable averages to jazz up the formula a bit and make it more suitable for application. In Munshi [2015c] and Munshi [2016], where I was considering the arithmetic twist aspect, it was more natural to take averages over q and ψ. One would imagine that when the problem in focus is in the t-aspect or spectral aspect, one would need to take average over k. Taking averages over nebentypus and level, executing the sum over ψ in the second part, and applying the reciprocity relations, we get

$$\delta(m-n) \approx \frac{1}{Q^2} \sum_{q \sim Q} \sum_{\substack{\psi \bmod q \\ \psi(-1)=(-1)^k}} \sum_{f \in H_k(q,\psi)} w_f^{-1} \lambda_f(n) \overline{\lambda_f(m)}$$

$$- \frac{2\pi i^{-k}}{Q} \sum_{q \sim Q} \Big[\frac{1}{N} \sum_{c \sim C} \sum_{\substack{a \bmod c \\ (a,c)=1}} e_c((a+1)\bar{q}m + (\bar{a}+1)\bar{q}n) \Big],$$

with $C = N/Q$. The first part of the formula involves $GL(2)$ harmonics, or Fourier coefficients of modular forms, the second part is like the usual circle method formula, and in fact matches in length to that if we take $Q = \sqrt{N}$. There are however two advantages - first one can take larger Q and thereby make c smaller, and secondly one can take advantage of the extra averaging over q. The cost one pays is of course the introduction of the more complicated $GL(2)$ harmonics in the formula. This version of delta method was introduced in Munshi [2015c] to tackle the M-aspect subconvexity for $L(s, \pi \otimes \chi)$ with π a $SL(3, \mathbb{Z})$ form. In the follow up paper Munshi [2016] a much simplified version of the approach was given. The second version of the proof is more in line with Munshi [2015d] and Munshi [2015b].

Theorem 2. *Let π be a Hecke-Maass cusp form for $SL(3, \mathbb{Z})$ and χ is a primitive character modulo M. Then we have*

$$L(\tfrac{1}{2}, \pi \otimes \chi) \ll M^{\frac{3}{4}-\theta},$$

for some explicitly computable $\theta > 0$.

The reader perhaps has already realised that the above method is robust enough and that these results should generalise to Hecke-Maass cusp forms for any congruence subgroups of $SL(3, \mathbb{Z})$. Also in the last stated theorem we can have a subconvex bound at any point on the critical line with polynomial dependence on t. The work of Blomer [2012] for quadratic twists of symmetric square lifts, in contrast, is only for the central point. There are other trace formulas which can be utilised in the same fashion. For example the Kuznetsov trace formula can be used to give an expansion of the delta involving Fourier coefficients of Maass forms and Einsenstein series. I believe that this would be the key in settling the weight/spectral aspect subconvexity for the symmetric square L-functions or the t-aspect subconvexity for $GL(4)$ L-functions. My recent preprint Munshi [2017b] addresses the level aspect subconvexity for the symmetric square L-function. Let me also mention that a preprint of Blomer and Buttcane [2015] gives a partial result towards settling the spectral aspect subconvexity for $GL(3)$ L-functions. They do not use the above method.

Few recent works have used the above method to get strong bounds for lower degree L-functions. First Aggarwal [2017] and Singh [2017] have independently revisited Good's problem using the method of Munshi [2015b]. Here it is likely that this method is strong enough to yield the Weyl exponent. I have shown that the $GL(2)$ delta method can be used to prove the Burgess exponent both for twists of $GL(2)$ forms and for classical Dirichlet L-functions Munshi [2017a]. New ideas are still required to break the Burgess barrier. However in a joint work with Singh Munshi and Singh [2017], we have shown that the Weyl bound holds in the M-aspect for twists of $GL(2)$ L-functions when the modulus is a suitable prime power, for example if $M = p^3$ for a prime p. This work again uses the congruence-equation trick. In this context, let me mention that Milićević [2016] and Blomer and Milićević [2015] have developed a p-adic version of the Van der Corput method, which yields a sub-Weyl bound for Dirichlet L-functions when the modulus is a suitably high p power, and gives a sub-Burgess bound (which asymptotically decreases to Weyl bound) for twists of $GL(2)$ L-functions by similar characters.

Finally I would like to mention that Holowinsky and Nelson [n.d.] have come up with a much 'abridged version' of the method. This have simplified and shortened the proofs substantially. Their work is pivoted on a crucial observation that there is a 'central identity' which makes the circle method approach to subconvexity work. They have shown that in many cases this central identity can be derived using simpler summation formulas, completely avoiding the circle method. However one drawback that still remains in their work is that it is not clear how to predict this central identity without taking recourse to the circle method approach. This is now an active topic of research. First Holowinsky-Nelson have given a simpler proof for the twists of $GL(3)$ L-functions. Y. Lin [n.d.] have used this approach to give a hybrid bound for $GL(3)$ L-functions, and Aggarwal, Holowinsky, Lin, and Sun [n.d.] have given the non-circle method version of Munshi [2017a]. Apart from the problem of breaking the longstanding Burgess barrier, the following two problems should be the main focus of the circle method approach.

Problem 1: Weight/spectral aspect subconvexity for symmetric square L-functions.

Problem 2: t-aspect subconvexity for $GL(4)$ L-functions.

I hope one of these would be solved before the next International Congress.

References

K. Aggarwal, R. Holowinsky, Y. Lin, and Q. Sun (n.d.). "Burgess bound for twisted L-functions via a trivial delta method" (cit. on p. 392).

Keshav Aggarwal (July 2017). "t-aspect subconvexity for $GL(2)$ L-functions". arXiv: 1707.07027 (cit. on p. 392).

Joseph Bernstein and Andre Reznikov (2010). "Subconvexity bounds for triple L-functions and representation theory". *Ann. of Math. (2)* 172.3, pp. 1679–1718. MR: 2726097 (cit. on p. 381).

Valentin Blomer (2012). "Subconvexity for twisted L-functions on $\mathrm{GL}(3)$". *Amer. J. Math.* 134.5, pp. 1385–1421. MR: 2975240 (cit. on pp. 388, 391).

Valentin Blomer and Jack Buttcane (Apr. 2015). "On the subconvexity problem for L-functions on GL(3)". arXiv: 1504.02667 (cit. on p. 391).

Valentin Blomer and Gergely Harcos (2008). "Hybrid bounds for twisted L-functions". *J. Reine Angew. Math.* 621, pp. 53–79. MR: 2431250 (cit. on pp. 381, 385).

Valentin Blomer and Djordje Milićević (2015). "p-adic analytic twists and strong subconvexity". *Ann. Sci. Éc. Norm. Supér. (4)* 48.3, pp. 561–605. MR: 3377053 (cit. on p. 392).

T. D. Browning and R. Munshi (2013). "Rational points on singular intersections of quadrics". *Compos. Math.* 149.9, pp. 1457–1494. MR: 3109729 (cit. on p. 388).

D. A. Burgess (1963). "On character sums and L-series. II". *Proc. London Math. Soc. (3)* 13, pp. 524–536. MR: 0148626 (cit. on pp. 385, 386).

V. A. Bykovskiĭ (1996). "A trace formula for the scalar product of Hecke series and its applications". *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 226.Anal. Teor. Chisel i Teor. Funktsiĭ. 13, pp. 14–36, 235–236. MR: 1433344.

J. B. Conrey and H. Iwaniec (2000). "The cubic moment of central values of automorphic L-functions". *Ann. of Math. (2)* 151.3, pp. 1175–1216. MR: 1779567 (cit. on p. 388).

W. Duke, J. B. Friedlander, and H. Iwaniec (1993). "Bounds for automorphic L-functions". *Invent. Math.* 112.1, pp. 1–8. MR: 1207474 (cit. on pp. 381, 385).

– (1994). "Bounds for automorphic L-functions. II". *Invent. Math.* 115.2, pp. 219–239. MR: 1258904 (cit. on p. 381).

– (1995). "Class group L-functions". *Duke Math. J.* 79.1, pp. 1–56. MR: 1340293 (cit. on p. 381).

– (2001). "Bounds for automorphic L-functions. III". *Invent. Math.* 143.2, pp. 221–248. MR: 1835388 (cit. on p. 381).

– (2002). "The subconvexity problem for Artin L-functions". *Invent. Math.* 149.3, pp. 489–577. MR: 1923476 (cit. on p. 381).

J. Friedlander and H. Iwaniec (1992). "A mean-value theorem for character sums". *Michigan Math. J.* 39.1, pp. 153–159. MR: 1137896 (cit. on p. 381).

John B. Friedlander (1995). "Bounds for L-functions". In: *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*. Birkhäuser, Basel, pp. 363–373. MR: 1403937 (cit. on pp. 381–383).

Anton Good (1982). "The square mean of Dirichlet series associated with cusp forms". *Mathematika* 29.2, 278–295 (1983). MR: 696884 (cit. on p. 384).

Gergely Harcos and Philippe Michel (2006). "The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points. II". *Invent. Math.* 163.3, pp. 581–655. MR: 2207235 (cit. on p. 381).

D. R. Heath-Brown (1980). "Hybrid bounds for Dirichlet L-functions. II". *Quart. J. Math. Oxford Ser. (2)* 31.122, pp. 157–167. MR: 576334 (cit. on p. 386).

R. Holowinsky and P. Nelson (n.d.). "Subconvex bounds on GL3 via degeneration to frequency zero" (cit. on p. 392).

H. Iwaniec and P. Sarnak (2000). "Perspectives on the analytic theory of L-functions". *Geom. Funct. Anal.* Special Volume, Part II. GAFA 2000 (Tel Aviv, 1999), pp. 705–741. MR: 1826269 (cit. on pp. 382, 383).

Henryk Iwaniec (1980). "Fourier coefficients of cusp forms and the Riemann zeta-function". In: *Seminar on Number Theory, 1979–1980 (French)*. Univ. Bordeaux I, Talence, Exp. No. 18, 36. MR: 604215 (cit. on p. 384).

E. Kowalski, P. Michel, and J. VanderKam (2002). "Rankin-Selberg L-functions in the level aspect". *Duke Math. J.* 114.1, pp. 123–191. MR: 1915038 (cit. on p. 381).

Yuk-Kam Lau, Jianya Liu, and Yangbo Ye (2006). "A new bound $k^{2/3+\epsilon}$ for Rankin-Selberg L-functions for Hecke congruence subgroups". *IMRP Int. Math. Res. Pap.* Art. ID 35090, 78. MR: 2235495 (cit. on p. 385).

Xiaoqing Li (2011). "Bounds for $\mathrm{GL}(3) \times \mathrm{GL}(2)$ L-functions and $\mathrm{GL}(3)$ L-functions". *Ann. of Math. (2)* 173.1, pp. 301–336. MR: 2753605 (cit. on pp. 388, 389).

Y. Lin (n.d.). "Bounds for twists of $GL(3)$ L-functions" (cit. on p. 392).

P. Michel (2004). "The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points". *Ann. of Math. (2)* 160.1, pp. 185–236. MR: 2119720 (cit. on p. 381).

Philippe Michel (2007). "Analytic number theory and families of automorphic L-functions". In: *Automorphic forms and applications*. Vol. 12. IAS/Park City Math. Ser. Amer. Math. Soc., Providence, RI, pp. 181–295. MR: 2331346 (cit. on pp. 382, 383).

Philippe Michel and Akshay Venkatesh (2006). "Equidistribution, L-functions and ergodic theory: on some problems of Yu. Linnik". In: *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, pp. 421–457. MR: 2275604 (cit. on pp. 381, 382).

– (2010). "The subconvexity problem for GL_2". *Publ. Math. Inst. Hautes Études Sci.* 111, pp. 171–271. MR: 2653249 (cit. on p. 381).

Djordje Milićević (2016). "Sub-Weyl subconvexity for Dirichlet L-functions to prime power moduli". *Compos. Math.* 152.4, pp. 825–875. MR: 3484115 (cit. on p. 392).

Ritabrata Munshi (2013a). "Bounds for twisted symmetric square L-functions". *J. Reine Angew. Math.* 682, pp. 65–88. MR: 3181499 (cit. on p. 388).

– (2013b). "Bounds for twisted symmetric square L-functions—III". *Adv. Math.* 235, pp. 74–91. MR: 3010051 (cit. on p. 388).

– (2014). "The circle method and bounds for L-functions—I". *Math. Ann.* 358.1-2, pp. 389–401. MR: 3158002 (cit. on p. 387).

– (2015a). "Pairs of quadrics in 11 variables". *Compos. Math.* 151.7, pp. 1189–1214. MR: 3371491 (cit. on p. 388).

– (2015b). "The circle method and bounds for L-functions—III: t-aspect subconvexity for $GL(3)$ L-functions". *J. Amer. Math. Soc.* 28.4, pp. 913–938. MR: 3369905 (cit. on pp. 389, 391, 392).

– (2015c). "The circle method and bounds for L-functions—IV: Subconvexity for twists of $\mathrm{GL}(3)$ L-functions". *Ann. of Math. (2)* 182.2, pp. 617–672. MR: 3418527 (cit. on pp. 390, 391).

– (2015d). "The circle method and bounds for L-functions, II: Subconvexity for twists of $\mathrm{GL}(3)$ L-functions". *Amer. J. Math.* 137.3, pp. 791–812. MR: 3357122 (cit. on pp. 388, 391).

– (Apr. 2016). "Twists of $GL(3)$ L-functions". arXiv: 1604.08000 (cit. on pp. 390, 391).

– (Oct. 2017a). "A note on Burgess bound". arXiv: 1710.02354 (cit. on p. 392).

– (Sept. 2017b). "Subconvexity for symmetric square L-functions". arXiv: 1709.05615 (cit. on p. 391).

Ritabrata Munshi and Saurabh Kumar Singh (June 2017). "Weyl bound for p-power twist of $GL(2)$ L-functions". arXiv: 1706.03985 (cit. on p. 392).

Peter Sarnak (1995). "Arithmetic quantum chaos". In: *The Schur lectures (1992) (Tel Aviv)*. Vol. 8. Israel Math. Conf. Proc. Bar-Ilan Univ., Ramat Gan, pp. 183–236. MR: 1321639 (cit. on p. 383).

– (1998). "L-functions". In: *Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998)*. Extra Vol. I, pp. 453–465. MR: 1648042 (cit. on p. 382).

– (2001). "Estimates for Rankin-Selberg L-functions and quantum unique ergodicity". *J. Funct. Anal.* 184.2, pp. 419–453. MR: 1851004 (cit. on p. 385).

Atle Selberg (1965). "On the estimation of Fourier coefficients of modular forms". In: *Proc. Sympos. Pure Math., Vol. VIII*. Amer. Math. Soc., Providence, R.I., pp. 1–15. MR: 0182610 (cit. on p. 384).

Saurabh Kumar Singh (June 2017). "t-aspect subconvexity bound for $GL(2)$ L-functions". arXiv: 1706.04977 (cit. on p. 392).

Akshay Venkatesh (2010). "Sparse equidistribution problems, period bounds and subconvexity". *Ann. of Math. (2)* 172.2, pp. 989–1094. MR: 2680486.

Han Wu (2014). "Burgess-like subconvex bounds for $\mathrm{GL}_2 \times \mathrm{GL}_1$". *Geom. Funct. Anal.* 24.3, pp. 968–1036. MR: 3213837 (cit. on p. 381).

Received 2017-11-27.

Ritabrata Munshi
School of Mathematics
Tata Institute of Fundamental Research
1 Dr. Homi Bhabha Road
Colaba, Mumbai 400005
India

and

Indian Statistical Institute
208 B.T. Road
Kolkata 700108
India
rmunshi@math.tifr.res.in

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (395–416)

ARITHMETIC MODELS FOR SHIMURA VARIETIES

Georgios Pappas

Abstract

We describe recent work on the construction of well-behaved arithmetic models for large classes of Shimura varieties and report on progress in the study of these models.

Introduction

Before the term became standard, certain "Shimura varieties" such as modular curves, Hilbert-Blumenthal varieties, and Siegel modular varieties, were already playing an important role in number theory. Indeed, these are, respectively, quotients of the domains on which modular, Hilbert, and Siegel modular forms, are defined. In a series of groundbreaking works Shimura [1964, 1970], Shimura initiated the arithmetic study of general quotients $\Gamma\backslash H$ of a hermitian symmetric domain H by the action of a discrete congruence arithmetic group Γ of holomorphic automorphisms of H. Such quotients are complex algebraic varieties and Shimura used the theory of moduli and of complex multiplication of abelian varieties to construct canonical models over explicit number fields for many of them.

Deligne reformulated and generalized Shimura's theory and emphasized the group and motivic theoretic source of the constructions (Deligne [1971, 1979] and Deligne, Milne, Ogus, and Shih [1982]). In Deligne's elegant definition, one starts with a pair $(\mathbb{G}, X)$ of a connected reductive algebraic group $\mathbb{G}$ defined over the rational numbers $\mathbb{Q}$ and a $\mathbb{G}(\mathbb{R})$-conjugacy class $X = \{h\}$ of an algebraic group homomorphism $h : \mathbb{S} \to \mathbb{G} \otimes_{\mathbb{Q}} \mathbb{R}$. Here, $\mathbb{S} = \mathrm{Res}_{\mathbb{C}/\mathbb{R}}(\mathbb{G}_{\mathrm{m}})$ is the algebraic torus over $\mathbb{R}$ whose real points are the group $\mathbb{C}^\times$. When the pair $(\mathbb{G}, X)$ satisfies Deligne's conditions (Deligne [1979, (2.1.1.1)-(2.1.1.3)]), we say that $(\mathbb{G}, X)$ is a *Shimura datum*. These conditions imply that each connected component

Partially supported by NSF grants DMS-1360733 and DMS-1701619.
MSC2010: primary 11G18; secondary 14G35.

X^+ of X is naturally a hermitian symmetric domain. The Shimura varieties $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$ are then the quotients

$$\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X) = \mathbb{G}(\mathbb{Q})\backslash X \times (\mathbb{G}(\mathbb{A}_f)/\mathbb{K})$$

for $\mathbb{K}$ an open compact subgroup of the finite adelic points $\mathbb{G}(\mathbb{A}_f)$ of $\mathbb{G}$. Here, $\mathbb{G}(\mathbb{Q})$ acts diagonally on $X \times (\mathbb{G}(\mathbb{A}_f)/\mathbb{K})$, with the action on X given by $\mathbb{G}(\mathbb{Q}) \subset \mathbb{G}(\mathbb{R})$ followed by conjugation, and on $\mathbb{G}(\mathbb{A}_f)/\mathbb{K}$ by $\mathbb{G}(\mathbb{Q}) \subset \mathbb{G}(\mathbb{A}_f)$ followed by left translation. Such a quotient is the disjoint union of a finite number of quotients of the form $\Gamma\backslash X^+$, where Γ are discrete congruence arithmetic groups. Hence, $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$ has a natural complex analytic structure induced from that on X. In fact, by work of Baily and Borel, there is a quasi-projective complex algebraic structure on $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$. Following the original work of Shimura and others, the existence of canonical models of $\mathrm{Sh}_K(\mathbb{G}, X)$ over a number field (the "reflex field" $\mathbb{E} = \mathbb{E}(\mathbb{G}, X)$) was shown in all generality by Borovoĭ [1987] and Milne (see Milne [1990] and the references there).

Many of the applications of Shimura varieties in number theory depend on understanding models of them over the ring of integers $\mathcal{O}_{\mathbb{E}}$ of the reflex field, or over localizations and completions of $\mathcal{O}_{\mathbb{E}}$. Indeed, perhaps the main application of Shimura varieties is to Langlands' program to associate Galois representations to automorphic representations. A related goal is to express the Hasse-Weil zeta functions of Shimura varieties as a product of automorphic L-functions.

After some earlier work by Eichler, Shimura, Sato and Ihara, a general plan for realizing this goal was given by Langlands, first for the local factor of the zeta function at a prime of good (*i.e.* smooth) reduction.[1] Langlands suggested expressing the numbers of points over finite fields of an integral model in terms of orbital integrals which appear in versions of the Arthur-Selberg trace formula. This was extended and realized in many cases, mainly by (Kottwitz [1990, 1992], see Langlands and Ramakrishnan [1992]). It is now often referred to as the Langlands-Kottwitz method. Langlands also considered the local factor of the zeta function for an example of a Shimura surface at a prime of bad (non-smooth) reduction Langlands [1979]. This example was treated carefully and the argument was extended to a larger class of Shimura varieties by Rapoport and Zink [1982] and by Rapoport [1990]. At primes of bad reduction, the singularities of the model have to be accounted for; the points need to be weighted by the semi-simple trace of Frobenius on the sheaf of nearby cycles. In fact, the Galois action on the nearby cycles can be used to study the local Langlands correspondence, as in the work of Harris and Taylor [2001]. In relation to this, Scholze recently extended the Langlands-Kottwitz method so that it can be applied, in principle at least, to the general case of bad reduction (Scholze [2013]).

[1] At least for proper Shimura varieties; in our brief report, we will ignore the issues arising from non-properness and the extensive body of work on compactifications.

A starting point of all the above is the existence of a reasonably well-behaved arithmetic model of the Shimura variety. For some time, such models could only be constructed for Shimura varieties of PEL type, *i.e.* those given as moduli spaces of abelian varieties with additional polarization, endomorphism, and level structures, and over primes at which the level subgroup is "parahoric" (*e.g.* Rapoport and Zink [1996]). Recently, due to advances in p-adic Hodge theory and in our understanding of the underlying group theory, the construction has been extended to most Shimura varieties of "abelian type" (at good reduction by Kisin [2010], see also earlier work of Vasiu [1999]; at general parahoric level in Kisin and Pappas [2015]). These Shimura varieties include most cases with $\mathbb{G}$ a classical group.

The construction and properties of these models is the subject of our report. There are, of course, more uses for these in number theory besides in the Langlands program. For example, one could mention showing Gross-Zagier type formulas via intersection theory over the integers, or developments in the theory of p-adic automorphic forms. Here, we view their construction and study as a topic of its own and discuss it independently of applications.

In fact, there are deep relations and analogies between this subject and the study of other spaces of interest in number theory, representation theory, and the geometric Langlands program, such as moduli spaces of bundles, or versions of affine Grassmannians and flag varieties. Increasingly, these connections, especially with the geometric side of Langlands program, are taking center stage. For example, certain p-adically integral models of homogeneous spaces that appear as subschemes of global ("Beilinson-Drinfeld") affine Grassmannians, the so-called "local models", play an important role. These local models also appear in the theory of deformations of Galois representations Kisin [2009]. After first giving some background, we start by discussing local models. We then describe results on arithmetic models of Shimura varieties and their reductions, and finish with an account of the local theory of Rapoport-Zink formal schemes.

Acknowledgments. I would like to thank my collaborators, especially M. Kisin and X. Zhu, for making much of the progress reported on here possible, and M. Rapoport for generously sharing his knowledge over the years.

1 Recollections on p-adic groups

Let G be a connected reductive group over the field of p-adic numbers $\mathbb{Q}_p$ for a prime number p. Let $\bar{\mathbb{Q}}_p$ be an algebraic closure of $\mathbb{Q}_p$. We denote by L the p-adic completion of the maximal unramified extension of $\mathbb{Q}_p$ in $\bar{\mathbb{Q}}_p$ and by $\mathcal{O}_L$ the integers of L. We will also denote by $k = \bar{\mathbb{F}}_p$ the algebraically closed residue field of L and by σ the automorphism of L which lifts the Frobenius $x \mapsto x^p$.

Recall that $g, g' \in G(L)$ are σ-conjugate, if there is $h \in G(L)$ with $g' = h^{-1} g \sigma(h)$. We denote by $B(G)$ the set of σ-conjugacy classes of $G(L)$. Recall the functorial surjective homomorphism $\kappa_G(L) : G(L) \to \pi_1(G)_I$ from Kottwitz [1997]. Here, $\mathrm{Gal}(\bar{L}/L) \simeq I \subset \mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$ is the inertia subgroup and $\pi_1(G)_I$ the inertia coinvariants of the algebraic fundamental group $\pi_1(G)$ of G over $\bar{\mathbb{Q}}_p$. We denote the kernel of $\kappa_G(L)$ by $G(L)_1$.

Let S be a maximal split torus of G_L. By Steinberg's theorem, G_L is quasi-split and the centralizer of S is a maximal torus T of G_L. Denote by N the normalizer of T. The quotient $\widetilde{W} = \widetilde{W}_{G,S} = N(L)/T(L)_1$ is the *Iwahori-Weyl group* associated to S. It is an extension of the relative Weyl group $W_0 = N(L)/T(L)$ by $\pi_1(T)_I$. Since $\pi_1(T) = X_*(T)$ (the group of cocharacters of T over $\bar{L}$), we obtain an exact sequence $1 \to X_*(T)_I \to \widetilde{W} \to W_0 \to 1$.

Suppose that $\{\mu\}$ is the conjugacy class of a cocharacter $\mu : \mathbb{G}_{\mathrm{m}\bar{\mathbb{Q}}_p} \to G_{\bar{\mathbb{Q}}_p}$. Then $\{\mu\}$ is defined over the *local reflex field* E which is a finite extension of $\mathbb{Q}_p$ contained in $\bar{E} = \bar{\mathbb{Q}}_p$. Denote by $\mathcal{O}_E$ the integers of E and by k_E its residue field. There is a corresponding homogeneous space $G_{\bar{E}}/P_{\mu^{-1}}$ which has a canonical model $X_\mu = X(\{\mu\})$ defined over E. Here, P_ν denotes the parabolic subgroup that corresponds to the cocharacter ν.

A pair $(G, \{\mu\})$ of a connected reductive group G over $\mathbb{Q}_p$, together with a conjugacy class $\{\mu\}$ of a cocharacter μ as above, is a *local Shimura pair*,[2] if μ is *minuscule* (*i.e.*, for any root α of $G_{\bar{\mathbb{Q}}_p}$, $\langle \alpha, \mu \rangle \in \{-1, 0, 1\}$.)

We denote by $\mathcal{B}(G, \mathbb{Q}_p)$ the (extended) Bruhat-Tits building of $G(\mathbb{Q}_p)$ Bruhat and Tits [1972, 1984] and Tits [1979]. The group $G(\mathbb{Q}_p)$ acts on $\mathcal{B}(G, \mathbb{Q}_p)$ on the left. If Ω is a subset of $\mathcal{B}(G, \mathbb{Q}_p)$, we write $G(\mathbb{Q}_p)_\Omega = \{g \in G(\mathbb{Q}_p) \mid g \cdot y = y,\ \text{for all } y \in \Omega\}$ for the pointwise fixer of Ω. Similarly, we have the subgroup $G(L)_\Omega$ of $G(L)$.

By the main result of Bruhat and Tits [1984], if Ω is bounded and contained in an apartment, there is a smooth affine group scheme $\mathcal{G}_\Omega$ over $\mathrm{Spec}\,(\mathbb{Z}_p)$ with generic fiber G and with $\mathcal{G}_\Omega(\mathcal{O}_L) = G(L)_\Omega$, which is uniquely characterized by these properties. By definition, the "connected fixer" $G(\mathbb{Q}_p)^\circ_\Omega$ is $\mathcal{G}^\circ_\Omega(\mathbb{Z}_p)$, where $\mathcal{G}^\circ_\Omega$ is the connected component of $\mathcal{G}_\Omega$. It is a subgroup of finite index in $G(\mathbb{Q}_p)_\Omega$. When $\Omega = \{x\}$ is a point, we simply write $G(\mathbb{Q}_p)_x$ and $G(\mathbb{Q}_p)^\circ_x$. If Ω is an open facet and $x \in \Omega$, then $G(\mathbb{Q}_p)^\circ_\Omega = G(\mathbb{Q}_p)^\circ_x$.

A *parahoric* subgroup of $G(\mathbb{Q}_p)$ is any subgroup which is the connected stabilizer $G(\mathbb{Q}_p)^\circ_x$ of some point x in $\mathcal{B}(G, \mathbb{Q}_p)$ as above. These are open compact subgroups of $G(\mathbb{Q}_p)$. We call $\mathcal{G}^\circ_x$ a "parahoric group scheme".

We now recall some more terms and useful facts (Tits [1979], Haines and Rapoport [2008]): A point $x \in \mathcal{B}(G, \mathbb{Q}_p)$ is *hyperspecial*, when $\mathcal{G}_x$ is reductive; then $\mathcal{G}_x = \mathcal{G}^\circ_x$. Hyperspecial points exist if and only if G is *unramified* over $\mathbb{Q}_p$, *i.e.* if G is quasi-split and splits over an unramified extension of $\mathbb{Q}_p$. When x is a point in an open alcove,

[2] Compare this with the term "local Shimura datum" used in Rapoport and Viehmann [2014]. A local Shimura datum $(G, \{\mu\}, [b])$ also includes the choice of a σ-conjugacy class $[b]$.

then the parahoric $G(\mathbb{Q}_p)^\circ_x$ is called an *Iwahori* subgroup. Iwahori subgroups exist for all G. Each maximal open compact subgroup of $G(\mathbb{Q}_p)$ contains a parahoric subgroup with finite index. If the group $\pi_1(G)_I$ has no torsion, then $G(\mathbb{Q}_p)_x = G(\mathbb{Q}_p)^\circ_x$, for all x. Then, all maximal open compact subgroups are parahoric.

Let $K = G(\mathbb{Q}_p)^\circ_x$ be parahoric and denote by $\tilde{K}$ the corresponding parahoric subgroup $G(L)^\circ_x$ of $G(L)$; we have $\tilde{K} = \mathcal{G}^\circ_x(\mathcal{O}_L)$.

Suppose x lies in the apartment associated to S. A choice of an alcove $C \subset \mathcal{B}(G, L)$ contained in that apartment provides the Iwahori-Weyl group $\widetilde{W} = \widetilde{W}_{G,S}$ with a Bruhat partial order $\leq$. Suppose $x \in C$ and set $\widetilde{W}^{\tilde{K}} = (N(L) \cap \tilde{K})/T(L)_1$, which is a subgroup of $\widetilde{W}$. The inclusion $N(L) \subset G(L)$ induces Haines and Rapoport [ibid.] a bijection

$$\widetilde{W}^{\tilde{K}}\backslash\widetilde{W}/\widetilde{W}^{\tilde{K}} \xrightarrow{\sim} \tilde{K}\backslash G(L)/\tilde{K}.$$

There is also a partial order $\leq$ on these double cosets: Set $[w] = \widetilde{W}^{\tilde{K}} w \widetilde{W}^{\tilde{K}}$. Then $[w_1] \leq [w_2]$ if and only if there are $w_1' \in [w_1]$, $w_2' \in [w_2]$, with $w_1' \leq w_2'$.

We refer the reader to Pappas, Rapoport, and Smithling [2013] for the definition of the $\{\mu\}$*-admissible subset*

$$\mathrm{Adm}_{\tilde{K}}(\{\mu\}) \subset \widetilde{W}^{\tilde{K}}\backslash\widetilde{W}/\widetilde{W}^{\tilde{K}} \simeq \tilde{K}\backslash G(L)/\tilde{K}$$

of Kottwitz and Rapoport. This is a finite set which has the following property: If $[w] \in \mathrm{Adm}_{\tilde{K}}(\{\mu\})$ and $[w'] \leq [w]$, then $[w'] \in \mathrm{Adm}_{\tilde{K}}(\{\mu\})$.

Let us continue with the above set-up. The *affine Deligne-Lusztig set* $X_K(\{\mu\}, b)$ associated to G, $\{\mu\}$, K, and $b \in G(L)$, is the subset of $G(L)/\tilde{K}$ consisting of those cosets $g\tilde{K}$ for which $g^{-1}b\sigma(g) \in \tilde{K}w\tilde{K}$, for some $[w] \in \mathrm{Adm}_{\tilde{K}}(\{\mu\})$.

If b' and b are σ-conjugate $b' = h^{-1}b\sigma(h)$, then $g\tilde{K} \mapsto hg\tilde{K}$ gives a bijection $X_K(\{\mu\}, b) \xrightarrow{\sim} X_K(\{\mu\}, b')$. The group $J_b(\mathbb{Q}_p) = \{j \in G(L) \mid j^{-1}b\sigma(j) = b\}$ acts on $X_K(\{\mu\}, b)$ by $j \cdot g\tilde{K} = jg\tilde{K}$. Set $f = [k_E : \mathbb{F}_p]$. The identity $\Phi_E(g\tilde{K}) = b\sigma(b)\cdots\sigma^{f-1}(b)\sigma^f(g)\tilde{K}$ defines a map $\Phi_E : X_K(\{\mu\}, b) \to X_K(\{\mu\}, b)$.

Our last reminder is of the affine Grassmannian of G. By definition, the *affine Grassmanian* Aff_G of G is the ind-projective ind-scheme over Spec $(\mathbb{Q}_p)$ which represents (*e.g.* Pappas and Rapoport [2008]) the fpqc sheaf associated to the quotient presheaf given by $R \mapsto G(R(\!(t)\!))/G(R[\![t]\!])$.

2 Local models

Let $(G, \{\mu\})$ be a local Shimura pair and K a parahoric subgroup of $G(\mathbb{Q}_p)$. Denote by $\mathcal{G}$ the corresponding parahoric group scheme over Spec $(\mathbb{Z}_p)$ so that $K = \mathcal{G}(\mathbb{Z}_p)$ and set again $\tilde{K} = \mathcal{G}(\mathcal{O}_L)$. A form of the following appears in Rapoport [2005].

Conjecture 2.1. *There exists a projective and flat scheme* $\mathrm{M}_K(G,\{\mu\})$ *over* $\mathrm{Spec}\,(\mathcal{O}_E)$ *which supports an action of* $\mathcal{G}\otimes_{\mathbb{Z}_p}\mathcal{O}_E$ *and is such that:*

(i) The generic fiber $\mathrm{M}_K(G,\{\mu\})\otimes_{\mathcal{O}_E} E$ *is* G_E*-isomorphic to* X_μ*,*

(ii) There is a $\tilde{K}$*-equivariant bijection between* $\mathrm{M}_K(G,\{\mu\})(k)$ *and*

$$\{g\tilde{K}\in G(L)/\tilde{K}\mid \tilde{K}g\tilde{K}\in \mathrm{Adm}_{\tilde{K}}(\{\mu\})\}\subset G(L)/\tilde{K}.$$

The scheme $\mathrm{M}_K(G,\{\mu\})$ is a *local model* associated to $(G,\{\mu\})$ and K.

We now consider the following "tameness" condition:

(T) The group G *splits over a tamely ramified extension of* $\mathbb{Q}_p$ *and* p *does not divide the order of the fundamental group* $\pi_1(G^{\mathrm{der}})$ *of the derived group.*

Under the assumption (T), Conjecture 2.1 is shown in Pappas and Zhu [2013]. The construction of the local models in Pappas and Zhu [ibid.] is as follows.

(I) We first construct a smooth affine group scheme $\underline{\mathcal{G}}$ over $\mathbb{Z}_p[u]$ which, among other properties, satisfies:

1) the restriction of $\underline{\mathcal{G}}$ over $\mathbb{Z}_p[u,u^{-1}]$ is reductive,

2) the base change $\underline{\mathcal{G}}\otimes_{\mathbb{Z}_p[u]}\mathbb{Z}_p$, given by $u\mapsto p$, is isomorphic to the parahoric group scheme $\mathcal{G}$.

For this, we use that $\mathbb{Z}_p[u^{\pm 1}]=\mathbb{Z}_p[u,u^{-1}]\to\mathbb{Q}_p$, given by $u\mapsto p$, identifies the étale fundamental group of $\mathbb{Z}_p[u^{\pm1}]$ with the tame quotient of the Galois group of $\mathbb{Q}_p$; our tameness hypothesis enters this way. We first obtain the restriction $\underline{\mathcal{G}}|_{\mathbb{Z}_p[u^{\pm1}]}$ from its constant Chevalley form $H\otimes_{\mathbb{Z}_p}\mathbb{Z}_p[u^{\pm1}]$ by using, roughly speaking, the "same twist" that gives G from its constant form $H\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$. This is quite straightforward when G is quasi-split and the twist is given using a diagram automorphism; the general case uses explicit Azumaya algebras over $\mathbb{Z}_p[u^{\pm1}]$. The extension $\underline{\mathcal{G}}$ to $\mathbb{Z}_p[u]$ of the reductive group scheme $\underline{\mathcal{G}}|_{\mathbb{Z}_p[u^{\pm1}]}$ is then given by generalizing some of the constructions in Bruhat and Tits [1984] to this two-dimensional set-up.

(II) Consider the functor that sends $\varphi:\mathbb{Z}_p[u]\to R$ to the set of isomorphism classes of pairs $(\mathcal{E},\beta)$, with $\mathcal{E}$ a $\underline{\mathcal{G}}$-torsor over $\mathbb{A}^1_R$ and β a section of $\mathcal{E}$ over $\mathrm{Spec}\,(R[u,(u-\varphi(u))^{-1}])$. We show that this is represented by an ind-projective ind-scheme (the *global* affine Grassmannian, *cf.* Beilinson and Drinfeld [1996]) $\mathrm{Aff}_{\underline{\mathcal{G}},\mathbb{A}^1}\to\mathbb{A}^1=\mathrm{Spec}\,(\mathbb{Z}_p[u])$. Using the construction of $\underline{\mathcal{G}}$ and Beauville and Laszlo [1994], we obtain an equivariant isomorphism

$$\mathrm{Aff}_G\xrightarrow{\sim}\mathrm{Aff}_{\underline{\mathcal{G}},\mathbb{A}^1}\otimes_{\mathbb{Z}_p[u]}\mathbb{Q}_p$$

where the base change is given by $u\mapsto p$. (The above isomorphism is induced by $t\mapsto u-p$, with the notation of Section 1 for Aff_G.)

(III) Since $\mathbb{G}_{\mathrm{m}}=\mathrm{Spec}\,(\mathbb{Q}_p[t,t^{-1}])$, our cocharacter μ defines $\mu(t)\in G(\bar{\mathbb{Q}}_p(\!(t)\!))$. In turn, this gives a $\bar{\mathbb{Q}}_p$-point $[\mu(t)]=\mu(t)G(\bar{\mathbb{Q}}_p[\![t]\!])$ of Aff_G.

Since μ is minuscule and the conjugacy class $\{\mu\}$ is defined over E, the orbit

$$G(\bar{\mathbb{Q}}_p[[t]]) \cdot [\mu(t)] \subset G(\bar{\mathbb{Q}}_p((t)))/G(\bar{\mathbb{Q}}_p[[t]]) = \mathrm{Aff}_G(\bar{\mathbb{Q}}_p),$$

is equal to the set of $\bar{\mathbb{Q}}_p$-points of a *closed* subvariety S_μ of $\mathrm{Aff}_{G,E} := \mathrm{Aff}_G \otimes_{\mathbb{Q}_p} E$. The stabilizer $H_{\mu^{-1}} := G(\bar{\mathbb{Q}}_p[[t]]) \cap \mu(t)G(\bar{\mathbb{Q}}_p[[t]])\mu(t)^{-1}$ of $[\mu(t)]$ is the inverse image of $P_{\mu^{-1}}(\bar{\mathbb{Q}}_p)$ under $G(\bar{\mathbb{Q}}_p[[t]]) \to G(\bar{\mathbb{Q}}_p)$ given by reduction modulo t. This gives a G_E-equivariant isomorphism $X_\mu \xrightarrow{\sim} S_\mu$ which we can use to identify X_μ with $S_\mu \subset \mathrm{Aff}_{G,E}$.

Definition 2.2. *We define* $\mathrm{M}_K(G, \{\mu\})$ *to be the (flat, projective) scheme over* $\mathrm{Spec}\,(\mathcal{O}_E)$ *given by the reduced Zariski closure of the image of*

$$X_\mu \xrightarrow{\sim} S_\mu \subset \mathrm{Aff}_{G,E} \xrightarrow{\sim} \mathrm{Aff}_{\underline{\mathcal{G}},\mathbb{A}^1} \otimes_{\mathbb{Z}_p[u]} E$$

in the ind-scheme $\mathrm{Aff}_{\underline{\mathcal{G}},\mathbb{A}^1} \otimes_{\mathbb{Z}_p[u]} \mathcal{O}_E$.

The next includes the main general facts about the structure of $\mathrm{M}_K(G, \{\mu\})$ and can be extracted from the results of Pappas and Zhu [2013].

Theorem 2.3. *Suppose that* (T) *holds. The scheme* $\mathrm{M}_K(G, \{\mu\})$ *of Definition 2.2 satisfies Conjecture 2.1. In addition:*

a) The scheme $\mathrm{M}_K(G, \{\mu\})$ *is normal.*

b) The geometric special fiber $\mathrm{M}_K(G, \{\mu\}) \otimes_{\mathcal{O}_E} k$ *is reduced and admits a* $\mathcal{G} \otimes_{\mathbb{Z}_p} k$*-orbit stratification by locally closed and smooth strata* $S_{[w]}$*, with* $S_{[w]}(k) \simeq \{g\tilde{K} \in G(L)/\tilde{K} \mid [w] = \tilde{K}g\tilde{K}\}$*, for each* $[w] \in \mathrm{Adm}_{\tilde{K}}(\{\mu\})$.

c) The closure $\overline{S}_{[w]}$ *of each stratum is normal and Cohen-Macaulay and equal to the union* $\bigcup_{[w'] \leq [w]} S_{[w']}$*, where* $\leq$ *is given by the Bruhat order.*

One main ingredient is the proof, by Zhu [2014], of the coherence conjecture of Pappas and Rapoport [2008]. The coherence conjecture is a certain numerical equality in the representation theory of (twisted) Kac-Moody groups. Its statement is, roughly speaking, independent of the characteristic, and so enough to show in the function field case where more tools are available.

Before Pappas and Zhu [2013], there have been various, often ad hoc, constructions of local models obtained by using variants of linked Grassmannians (Rapoport and Zink [1996], Pappas [2000]; see the survey Pappas, Rapoport, and Smithling [2013] and the references within, and Section 3 below for an example).

An extension of the above construction of local models, to the case G is the restriction of scalars of a tame group from a wild extension, is given in Levin [2016]. One would expect that a completely general construction of the local model $\mathrm{M}_K(G, \{\mu\})$ can be given by using a hypothetical version of the affine Grassmannian in which $R \mapsto \mathcal{G}(R[[u]])$ is replaced by the functor $R \mapsto \mathcal{G}(W(R))$. Here $W(R)$ is the ring of p-typical Witt vectors of

R. (See Section 5 for a related construction when $\mathcal{G}$ is reductive.) Such Witt affine Grassmannians are defined in characteristic p by Zhu [2017] (but only "after perfection") and further studied by Bhatt and Scholze [2017]; other variations have been used by Scholze.

3 Global theory: Arithmetic models

Let $(\mathbb{G}, \{h\})$ be a Shimura datum as in the introduction. Define $\mu_h : \mathbb{G}_{m\mathbb{C}} \to \mathbb{G}_{\mathbb{C}}$ by $\mu_h(z) = h_{\mathbb{C}}(z, 1)$. The $G(\mathbb{C})$-conjugacy class $\{\mu_h\}$ is defined over the reflex field $\mathbb{E} \subset \bar{\mathbb{Q}} \subset \mathbb{C}$. Write $\mathbb{Z}_s$ for the maximal subtorus of the center $Z(\mathbb{G})$ which is $\mathbb{R}$-split but which has no $\mathbb{Q}$-split subtorus and set $\mathbb{G}^c = \mathbb{G}/\mathbb{Z}_s$.

Consider a $\mathbb{Q}$-vector space $\mathbb{V}$ with a perfect alternating bilinear pairing $\psi : \mathbb{V}\times\mathbb{V} \to \mathbb{Q}$. Set $\dim_{\mathbb{Q}}(\mathbb{V}) = 2g$. Let $\mathbb{G} = \mathrm{GSp}(\mathbb{V}, \psi)$ be the group of symplectic similitudes and $X = S^{\pm}$ the Siegel double space. This is the set of homomorphisms $h : \mathbb{S} \to \mathbb{G}_{\mathbb{R}}$ such that: (1) The $\mathbb{C}^{\times}$-action given by $h(\mathbb{R}) : \mathbb{C}^* \to \mathbb{G}(\mathbb{R})$ gives on $\mathbb{V}_{\mathbb{R}}$ a Hodge structure of type $\{(-1, 0), (0, -1)\}$: $\mathbb{V}_{\mathbb{C}} = \mathbb{V}^{-1.0} \oplus \mathbb{V}^{0,-1}$, and, (2) the form $(x, y) \mapsto \psi(x, h(i)y)$ on $\mathbb{V}_{\mathbb{R}}$ is positive (or negative) definite. The pairs $(\mathrm{GSp}(\mathbb{V}, \psi), S^{\pm})$ give the most important examples of Shimura data. By Riemann's theorem, if $\mathbb{V}_{\mathbb{Z}}$ is a $\mathbb{Z}$-lattice in $\mathbb{V}$, and $h \in S^{\pm}$, the quotient torus $\mathbb{V}^{-1,0}/\mathbb{V}_{\mathbb{Z}}$ is a complex abelian variety of dimension g. This leads to an interpretation of the Shimura varieties $\mathrm{Sh}_{\mathbb{K}}(\mathrm{GSp}(\mathbb{V}, \psi), S^{\pm})$ as moduli spaces of (polarized) abelian varieties of dimension g with level structure.

A Shimura datum $(\mathbb{G}, X)$ with $X = \{h\}$ is of *Hodge type*, when there is a symplectic space $(\mathbb{V}, \psi)$ over $\mathbb{Q}$ and a closed embedding $\rho : \mathbb{G} \hookrightarrow \mathrm{GSp}(\mathbb{V}, \psi)$ such that the composition $\rho \circ h$ lies in the Siegel double space $S^{\pm}$. Then, we also have $\mathbb{G} = \mathbb{G}^c$. In this case, the Shimura varieties $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$ parametrize abelian varieties together with (absolute) Hodge cycles (see below). A special class of Shimura data of Hodge type are those of *PEL type* Kottwitz [1992, §5], Rapoport and Zink [1996, Ch. 6]. For those, the corresponding Shimura varieties $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$ are (essentially) moduli schemes of abelian varieties together with polarization, endomorphisms and level structure.

A Shimura datum $(\mathbb{G}, X)$ is of *abelian type* if there is a datum of Hodge type $(\mathbb{G}_1, X_1)$ and a central isogeny $\mathbb{G}_1^{\mathrm{der}} \to \mathbb{G}^{\mathrm{der}}$ between derived groups which induces an isomorphism $(\mathbb{G}_1^{\mathrm{ad}}, X_1^{\mathrm{ad}}) \xrightarrow{\sim} (\mathbb{G}^{\mathrm{ad}}, X^{\mathrm{ad}})$. Here, the superscript $^{\mathrm{ad}}$ denotes passing to the adjoint group: If $X = \{h\}$, then $X^{\mathrm{ad}} = \{h^{\mathrm{ad}}\}$ with h^{ad} the composition of $h : \mathbb{S} \to \mathbb{G}_{\mathbb{R}}$ with $\mathbb{G}_{\mathbb{R}} \to \mathbb{G}_{\mathbb{R}}^{\mathrm{ad}}$. Roughly speaking, most Shimura data $(\mathbb{G}, X)$ with $\mathbb{G}$ a classical group are of abelian type.

Fix a prime number p and a prime $\mathfrak{p}$ of the reflex field $\mathbb{E}$ above p which is obtained from an embedding $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_p$. Let E be the completion of $\mathbb{E}$ at $\mathfrak{p}$ and set $G = \mathbb{G}_{\mathbb{Q}_p}$. Then $\{\mu_h\}$ gives a conjugacy class $\{\mu\}$ of G defined over the local reflex field E and $(G, \{\mu\})$ is a local Shimura pair.

Fix a *parahoric* subgroup $K = K_p \subset \mathbb{G}(\mathbb{Q}_p) = G(\mathbb{Q}_p)$. For any open compact subgroup $K^p \subset \mathbb{G}(\mathbb{A}_f^p)$ we can consider the Shimura variety $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$ over $\mathbb{E}$, where $\mathbb{K} = K_p K^p \subset \mathbb{G}(\mathbb{A}_f)$.

We now assume that a local model $\mathrm{M}_K(G, \{\mu\})$ as in Conjecture 2.1 is given for $(G, \{\mu\})$ and parahoric K.

Conjecture 3.1. *There is a scheme $\mathcal{S}_{K_p}(\mathbb{G}, X)$ over* $\mathrm{Spec}\,(\mathcal{O}_E)$ *which supports a right action of $\mathbb{G}(\mathbb{A}_f^p)$ and has the following properties:*

a) Any sufficiently small open compact subgroup $K^p \subset \mathbb{G}(\mathbb{A}_f^p)$ acts freely on $\mathcal{S}_{K_p}(\mathbb{G}, X)$, and the quotient $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X) := \mathcal{S}_{K_p}(\mathbb{G}, X)/K^p$ is a scheme of finite type over $\mathcal{O}_E$ which extends $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X) \otimes_{\mathbb{E}} E$. We have $\mathcal{S}_{K_p}(\mathbb{G}, X) = \varprojlim_{K^p} \mathcal{S}_{K_p K^p}(\mathbb{G}, X)$ where the limit is over all such $K^p \subset \mathbb{G}(\mathbb{A}_f^p)$.

b) For any discrete valuation ring $R \supset \mathcal{O}_E$ of mixed characteristic $(0, p)$, the map $\mathcal{S}_{K_p}(\mathbb{G}, X)(R) \to \mathcal{S}_{K_p}(\mathbb{G}, X)(R[1/p])$ is a bijection.

c) There is a smooth morphism of stacks over $\mathrm{Spec}\,(\mathcal{O}_E)$

$$\lambda : \mathcal{S}_{K_p}(\mathbb{G}, X) \to [(\mathcal{G}^c \otimes_{\mathbb{Z}_p} \mathcal{O}_E)\backslash \mathrm{M}_K(G, \{\mu\})]$$

which is invariant for the $\mathbb{G}(\mathbb{A}_f^p)$-action on the source and is such that the base change λ_E is given by the canonical principal $\mathbb{G}_E^c$-bundle over $\mathrm{Sh}_{K_p}(\mathbb{G}, X) \otimes_{\mathbb{E}} E$ (Milne [1990, III, §3]). Here, we set $\mathcal{G}^c = \mathcal{G}/\mathcal{Z}_s$, where $\mathcal{Z}_s$ is the Zariski closure of the central torus $Z_s \subset G$ in $\mathcal{G}$.

It is important to record that the existence of the smooth λ as in (c) implies:

(c') For each closed point $x \in \mathcal{S}_{K_p}(\mathbb{G}, X)$, there is a closed point $y \in \mathrm{M}_K(G, \{\mu\})$ and étale neighborhoods $U_x \to \mathcal{S}_{K_p}(\mathbb{G}, X)$ of x and $V_y \to \mathrm{M}_K(G, \{\mu\})$ of y, which are isomorphic over $\mathcal{O}_E$.

The significance of (c') is that the singularities of $\mathrm{M}_K(G, \{\mu\})$ control the singularities of $\mathcal{S}_{K_p}(\mathbb{G}, X)$, and so, by (a), also of the integral models $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$ for $\mathbb{K} = K_p K^p$, with K^p sufficiently small. In what follows, we will often assume that K^p is sufficiently small without explicitly saying so.

The "extension property" (b) ensures that the special fiber of $\mathcal{S}_{K_p}(\mathbb{G}, X)$ is sufficiently large, in particular, it cannot be empty. Unfortunately, it is not clear that properties (a)-(c) uniquely characterize the schemes $\mathcal{S}_{K_p}(\mathbb{G}, X)$. We will return to this question later.

Before we give some general results we discuss the illustrative example of the Siegel Shimura datum $(\mathrm{GSp}(\mathbb{V}, \psi), S^{\pm})$ as in Section 3.

Fix a prime p. Set $V = \mathbb{V}_{\mathbb{Q}_p}$ and denote the induced form on V also by ψ. For a $\mathbb{Z}_p$-lattice Λ in V, its dual is $\Lambda^\vee = \{x \in V \mid \psi(x, y) \in \mathbb{Z}_p\,,\text{for all } y \in \Lambda\}$. Now choose a chain of lattices $\{\Lambda_i\}_{i \in \mathbb{Z}}$ in V with the following properties: $\Lambda_i \subset \Lambda_{i+1}$ and $\Lambda_{i-m} = p\Lambda_i$, for all i and a fixed "period" $m > 0$, and $\Lambda_i^\vee = \Lambda_{-i+a}$, for all i and (fixed) $a = 0$ or 1. The chain $\{\Lambda_i\}_{i\in\mathbb{Z}}$ gives rise to a point in the building of $G = \mathrm{GSp}(V, \psi)$ over $\mathbb{Q}_p$. The corresponding parahoric K is the subgroup of all $g \in \mathrm{GSp}(V, \psi)$ such that $g\Lambda_i = \Lambda_i$, for all i. Every parahoric subgroup of $G(\mathbb{Q}_p)$ is obtained from such a lattice chain.

In this case, schemes $\mathcal{S}_K(\mathrm{GSp}(\mathbb{V}, \psi), S^\pm)$ that satisfy the conjecture are given by moduli spaces as we will now explain. Recall that the category $AV_{(p)}$ of abelian schemes up to prime-to-p isogeny has objects abelian schemes and morphisms given by tensoring the Hom groups by $\mathbb{Z}_{(p)}$. We consider the functor which associates to a $\mathbb{Z}_p$-algebra R the set of isomorphism classes of m-tuples $\{(A_i, \alpha_i, \tilde{\nu}, \eta)\}_{i\in\mathbb{Z}/m\mathbb{Z}}$ where (see Rapoport and Zink [1996, Ch. 6] for details):

1) A_i are up to prime-to-p isogeny projective abelian schemes over R,

2) $\alpha_i : A_i \to A_{i+1}$ are isogenies of height $\log_p([\Lambda_{i+1} : \Lambda_i])$ in $AV_{(p)}$ such that the compositions $\alpha_{r+m-1}\cdots\alpha_{r+1}\alpha_r = p$, for all $0 \le r < m$,

3) $\tilde{\nu}$ is a set of isomorphisms $\{\nu_i : A_i \xrightarrow{\sim} A^\vee_{-i+a}\}_{i\in\mathbb{Z}/m\mathbb{Z}}$ in $AV_{(p)}$ which are compatible with α_i, and is a $\mathbb{Z}_{(p)}^\times$-homogeneous polarization,

4) η is a prime-to-p level structure of type K^p on $\{(A_i, \alpha_i, \tilde{\nu})\}_i$.

When $K^p \subset \mathbb{G}(\mathbb{A}_f^p)$ is sufficiently small, this functor is represented by a scheme $\mathcal{S}_{KK^p}(\mathrm{GSp}(\mathbb{V}, \psi), S^\pm)$ over $\mathbb{Z}_p$. The limit over K^p is easily seen to satisfy (a) while (b) follows from the Néron-Ogg-Shafarevich criterion for good reduction of abelian varieties. Next, we discuss (c).

For a $\mathbb{Z}_p$-algebra R, set $\Lambda_{i,R} := \Lambda_i \otimes_{\mathbb{Z}_p} R$ and denote by $a_{i,R} : \Lambda_{i,R} \to \Lambda_{i+1,R}$ the map induced by the inclusion $\Lambda_i \to \Lambda_{i+1}$ and by $(\ ,\ )_i : \Lambda_{i,R} \times \Lambda_{-i+a,R} \to R$ the perfect R-bilinear pairing induced by ψ since $\Lambda_i^\vee = \Lambda_{-i+a}$.

A result of Görtz [2003] implies that, in this case, the local model $\mathrm{M}_K(G, \{\mu\})$ of Pappas and Zhu [2013] represents the functor which sends a $\mathbb{Z}_p$-algebra R to the set of sequences $\{\mathfrak{F}_i\}_{i\in\mathbb{Z}}$, where: $\mathfrak{F}_i \subset \Lambda_{i,R}$ is an R-submodule which is Zariski locally a direct summand of $\Lambda_{i,R}$ of rank g, and, for all i, we have: $a_i(\mathfrak{F}_i) \subset \mathfrak{F}_{i+1}$, $(\mathfrak{F}_i, \mathfrak{F}_{-i+a})_i = 0$, and $\mathfrak{F}_{i-m} = p\mathfrak{F}_i \subset p\Lambda_{i,R} = \Lambda_{i-m,R}$.

The morphism λ is defined by first showing that the "crystalline" system of modules, homomorphisms and pairings, obtained by applying the Dieudonné functor to $\{(A_i, \alpha_i, \nu_i)\}$, is locally isomorphic to the corresponding "Betti" system given by the lattice chain $\{\Lambda_i\}$ and ψ. Then, λ sends $\{(A_i, \alpha_i, \tilde{\nu}, \eta)\}$ to $\mathfrak{F}_i$ given by the Hodge filtration of A_i. It is smooth since, by Grothendieck-Messing theory, deformations of an abelian variety are determined by lifts of its Hodge filtration. This argument first appeared in de Jong [1993] and Deligne and Pappas [1994]. It extends to most PEL type cases Rapoport and Zink [1996]. In general, we need a different approach which we explain next.

We now discuss general results. Assume that p is odd and that (T) holds. Assume also that the local models $\mathrm{M}_K(G, \{\mu\})$ are as defined in Pappas and Zhu [2013]. The next result is shown in Kisin and Pappas [2015], following earlier work of Kisin [2010].

Theorem 3.2. *(i) Assume that $(\mathbb{G}, X)$ is of abelian type. Then Conjecture 3.1 with (c) replaced by (c'), is true for $(\mathbb{G}, X)$ and K.*

(ii) Assume that $(\mathbb{G}, X)$ is of Hodge type and that the parahoric subgroup K is equal to the stabilizer of a point in the Bruhat-Tits building of G. Then Conjecture 3.1 is true for $(\mathbb{G}, X)$ and K.

Let us try to explain some ideas in the proofs.

We will first discuss (ii). So, the Shimura datum $(\mathbb{G}, X)$ is of Hodge type and the parahoric subgroup $K \subset G(\mathbb{Q}_p)$ is the stabilizer of a point z in the building $\mathcal{B}(G, \mathbb{Q}_p)$, *i.e.* the stabilizer group scheme is already connected, $\mathcal{G} = \mathcal{G}_z = \mathcal{G}_z^\circ$.

After adjusting the symplectic representation $\rho : \mathbb{G} \hookrightarrow \mathrm{GSp}(\mathbb{V}, \psi)$ which gives the Hodge embedding, we can find a $\mathbb{Z}_p$-lattice $\Lambda \subset V = \mathbb{V}_{\mathbb{Q}_p}$ on which ψ takes $\mathbb{Z}_p$-integral values, such that ρ induces

1) a closed group scheme embedding $\rho_{\mathbb{Z}_p} : \mathcal{G} = \mathcal{G}_z \hookrightarrow \mathrm{GL}(\Lambda)$, and

2) an equivariant closed embedding $\rho_* : \mathrm{M}_K(G, \{\mu\}) \hookrightarrow \mathrm{Gr}(g, \Lambda^\vee) \otimes_{\mathbb{Z}_p} \mathcal{O}_E$.

Here $\mathrm{Gr}(g, \Lambda^\vee)$ is the Grassmannian over $\mathbb{Z}_p$ with R-points given by R-submodules $\mathcal{F} \subset \Lambda_R^\vee$ which are locally direct summands of rank $g = \dim_{\mathbb{Q}}(\mathbb{V})/2$. Finding Λ is subtle and uses that the representation ρ is *minuscule*.

Choose a $\mathbb{Z}$-lattice $\mathbb{V}_{\mathbb{Z}} \subset \mathbb{V}$ such that $\mathbb{V}_{\mathbb{Z}_{(p)}} := \mathbb{V}_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)} = \mathbb{V} \cap \Lambda$. We can now find $\{s_\alpha\}_\alpha \subset \mathbb{V}^{\otimes}_{\mathbb{Z}_{(p)}} \subset \mathbb{V}^{\otimes}$ such that the functor

$$R \mapsto \mathbb{G}_{\mathbb{Z}_{(p)}}(R) = \{g \in \mathrm{GL}(\mathbb{V}_{\mathbb{Z}_{(p)}} \otimes_{\mathbb{Z}_{(p)}} R) \mid g \cdot s_\alpha = s_\alpha, \text{ for all } \alpha\}$$

gives the unique flat group scheme $\mathbb{G}_{\mathbb{Z}_{(p)}}$ over $\mathbb{Z}_{(p)}$ that extends both $\mathbb{G}$ and $\mathcal{G}$. Here, $\mathbb{V}^{\otimes} := \bigoplus_{r,s \geq 0}(V^{\otimes r} \otimes (V^\vee)^{\otimes s})$, similarly for $\mathbb{V}^{\otimes}_{\mathbb{Z}_{(p)}}$.

Now fix $K^p \subset \mathbb{G}(\mathbb{A}_f^p)$ small enough so, in particular, $V_{\hat{\mathbb{Z}}} := V_{\mathbb{Z}} \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$ is stable by the action of K^p; then $V_{\hat{\mathbb{Z}}}$ is $\mathbb{K} = KK^p$-stable also. We can find $K'^p \subset \mathrm{GSp}(\mathbb{V})(\mathbb{A}_f^p)$ such that ρ gives a closed embedding

$$\iota : \mathrm{Sh}_{KK^p}(\mathbb{G}, X) \hookrightarrow \mathrm{Sh}_{K'_pK'^p}(\mathrm{GSp}(\mathbb{V}), S^\pm) \otimes_{\mathbb{Q}} \mathbb{E},$$

where K'_p is the subgroup of $\mathrm{GSp}(V, \psi)$ that stabilizes the lattice Λ, and such that $\mathbb{V}_{\hat{\mathbb{Z}}}$ is also stable by the action of $\mathbb{K}' = K'_pK'^p$.

We define $\mathcal{S}_{KK^p}(\mathbb{G}, X)$ to be the normalization of the reduced Zariski closure of the image of ι_E in the integral model $\mathcal{S}_{K'_pK'^p}(\mathrm{GSp}(\mathbb{V}), S^\pm) \otimes_{\mathbb{Z}_p} \mathcal{O}_E$ (which is a moduli scheme, as in Section 3) and set $\mathcal{S}_K(\mathbb{G}, X) := \varprojlim_{K^p} \mathcal{S}_{KK^p}(\mathbb{G}, X)$. Checking property

(a) is straightforward and (b) follows from the Néron-Ogg-Shafarevich criterion for good reduction. The hard work is in showing (c).

Since $\mathbb{V}_{\hat{\mathbb{Z}}}$ is stable by the action of $\mathbb{K}'$, which is sufficiently small, we have a universal abelian scheme over $\mathrm{Sh}_{\mathbb{K}'}(\mathrm{GSp}(\mathbb{V}), S^{\pm})$ which we can restrict via ι to obtain an abelian scheme $\mathcal{A}$ over $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)$. By construction, the tensors s_α above give sections $s_{\alpha,\mathrm{B}}$ of the local system of $\mathbb{Q}$-vector spaces over $\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)(\mathbb{C})$ with fibers $\mathrm{H}^1_{\mathrm{B}}(\mathcal{A}_v(\mathbb{C}), \mathbb{Q})^{\otimes}$ given using the first Betti (singular) cohomology of the fibers $\mathcal{A}_v(\mathbb{C})$ of the universal abelian variety. These are "Hodge cycles" in the sense that they are of type $(0, 0)$ for the induced Hodge structure on $\mathrm{H}^1_{\mathrm{B}}(\mathcal{A}(\mathbb{C}), \mathbb{Q})^{\otimes}$. We are going to chase these around using various comparison isomorphisms.

Let $\kappa \supset \mathbb{E}$ be a field with an embedding $\sigma : \bar{\kappa} \hookrightarrow \mathbb{C}$ over $\mathbb{E}$ of its algebraic closure. Suppose $x \in \mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X)(\kappa)$ and let $\mathcal{A}_x$ be the corresponding abelian variety over κ. There are comparison isomorphisms $\mathrm{H}^1_{B}(\mathcal{A}_x(\mathbb{C}), \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} \simeq \mathrm{H}^1_{\mathrm{dR}}(\mathcal{A}_x) \otimes_{\kappa,\sigma} \mathbb{C}$, and $\mathrm{H}^1_{B}(\mathcal{A}_x(\mathbb{C}), \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_l \simeq \mathrm{H}^1_{\mathrm{et}}(\mathcal{A}_x \otimes_\kappa \bar{\kappa}, \mathbb{Q}_l)$, for any prime l. Set $s_{\alpha,B,x}$ be the fiber of $s_{\alpha,B}$ over the complex point $\sigma(x)$ and denote by $s_{\alpha,\mathrm{dR},x}$, $s_{\alpha,\mathrm{et},x}$, the images of $s_{\alpha,B,x}$ under these isomorphisms. The tensors $s_{\alpha,\mathrm{et},x}$ are independent of the embedding σ, are fixed by the action of $\mathrm{Gal}(\bar{\kappa}/\kappa)$, and it follows from Deligne's "Hodge implies absolute Hodge" theorem that $s_{\alpha,\mathrm{dR},x}$ lie in $\mathrm{H}^1_{\mathrm{dR}}(\mathcal{A}_x)^{\otimes}$ and, in fact, in $F^0(\mathrm{H}^1_{\mathrm{dR}}(\mathcal{A}_x)^{\otimes})$ (Kisin [2010]).

For $l = p$, the corresponding $\mathrm{Gal}(\bar{\kappa}/\kappa)$-invariant tensors ("Tate cycles") $s_{\alpha,\mathrm{et},x}$ lie in the lattice $\mathrm{H}^1_{\mathrm{et}}(\mathcal{A}_x \otimes_\kappa \bar{\kappa}, \mathbb{Z}_p)^{\otimes} = T_p(\mathcal{A}_x)^{\otimes}$, where $T_p(\mathcal{A}_x)$ is the p-adic Tate module $\varprojlim_n (\mathcal{A}_x \otimes_\kappa \bar{\kappa})[p^n]$. In order to control the local structure of $\mathcal{S}_{KK^p}(\mathbb{G}, X)$ and eventually relate it to $\mathrm{M}_K(G, \{\mu\})$, we need to employ some form of crystalline deformation theory and so we also have to understand the "crystalline realization" of our tensors. Assume that $\kappa = F$ is a finite extension of E and that the abelian variety $\mathcal{A}_x$ has good reduction. Then we obtain an $\mathcal{O}_F$-valued point $\tilde{x}$ of $\mathcal{S}_{KK^p}(\mathbb{G}, X)$ that extends x and reduces to $\bar{x}$. A key point is to show that the Tate cycles $s_{\alpha,\mathrm{et},x}$ for $l = p$ give, via the étale/crystalline comparison, corresponding "nice" *integral* crystalline cycles $s_{\alpha,\mathrm{cris},\bar{x}} \in \mathbb{D}(\mathcal{A}_{\bar{x}})^{\otimes}$ on the Dieudonné module $\mathbb{D}(\mathcal{A}_{\bar{x}})$ of the abelian variety $\mathcal{A}_{\bar{x}}$ over the residue field of F. For this we need a suitably functorial construction that relates crystalline p-adic Galois representations of $\mathrm{Gal}(\bar{F}/F)$ to Frobenius semilinear objects integrally, and without restriction on the absolute ramification of F. This is provided by the theory of Breuil-Kisin modules Kisin [2009, 2010].

Let F_0 be the maximal unramified extension of $\mathbb{Q}_p$ contained in F. Denote by W_0 the ring of integers of F_0 and by $\varphi : W_0[\![u]\!] \to W_0[\![u]\!]$ the lift of Frobenius with $\varphi(u) = u^p$. Choose a uniformizer π of F which is a root of an Eisenstein polynomial $E(u) \in W_0[u]$. A Breuil-Kisin module $(\mathfrak{M}, \Phi)$ is a finite free $W_0[\![u]\!]$-module $\mathfrak{M}$ with an isomorphism $\Phi : \varphi^*(\mathfrak{M})[1/E(u)] \xrightarrow{\sim} \mathfrak{M}[1/E(u)]$, where $\varphi^*(\mathfrak{M}) := W_0[\![u]\!] \otimes_{\varphi, W_0[\![u]\!]} \mathfrak{M}$. Kisin has constructed a fully faithful tensor functor $T \mapsto \mathfrak{M}(T)$ from the category of $\mathrm{Gal}(\bar{F}/F)$-stable $\mathbb{Z}_p$-lattices in crystalline $\mathbb{Q}_p$-representations to the category of Breuil-Kisin modules.

Now let $T^\vee = \mathrm{Hom}_{\mathbb{Z}_p}(T, \mathbb{Z}_p)$ be the linear dual of the p-adic Tate module $T = T_p(\mathcal{Q}_x)$. Then there are natural isomorphisms Kisin [2009]

$$\varphi^*(\mathfrak{M}(T^\vee)/u\mathfrak{M}(T^\vee)) \simeq \mathbb{D}(\mathcal{Q}_{\bar{x}}), \quad \varphi^*(\mathfrak{M}(T^\vee)) \otimes_{W_0[\![u]\!], u\mapsto\pi} \mathcal{O}_F \simeq \mathrm{H}^1_{\mathrm{dR}}(\mathcal{Q}_{\tilde{x}}).$$

Applying Kisin's functor to the Tate cycles $s_{\alpha,\mathrm{et},x}$ gives Frobenius invariant tensors $\tilde{s}_\alpha \in \mathfrak{M}(T_p(\mathcal{Q}_x)^\vee)^\otimes$. By using the isomorphism above, these give the crystalline cycles $s_{\alpha,\mathrm{cris},\bar{x}} \in \mathbb{D}(\mathcal{Q}_{\bar{x}})^\otimes$. The main result now is:

Theorem 3.3. *There is a $W_0[\![u]\!]$-linear isomorphism*

$$\beta : \Lambda^\vee \otimes_{\mathbb{Z}_p} W_0[\![u]\!] \xrightarrow{\sim} \mathfrak{M}(T_p(\mathcal{Q}_x)^\vee)$$

such that $\beta^\otimes$ takes $s_\alpha \otimes 1$ to $\tilde{s}_\alpha$, for all α.

In addition to the properties of Kisin's functor, the important input in the proof is the statement that all $\mathcal{G}$-torsors over $\mathrm{Spec}\,(W_0[\![u]\!]) - \{(0,0)\}$ are trivial. This uses crucially that $\mathcal{G} = \mathcal{G}_z^\circ$ is parahoric.

Using β and the above, we obtain isomorphisms

$$\beta_{\mathrm{dR}} : \Lambda^\vee \otimes_{\mathbb{Z}_p} \mathcal{O}_F \xrightarrow{\sim} \mathrm{H}^1_{\mathrm{dR}}(\mathcal{Q}_{\tilde{x}}), \quad \beta_{\mathrm{cris}} : \Lambda^\vee \otimes_{\mathbb{Z}_p} W_0 \xrightarrow{\sim} \mathbb{D}(\mathcal{Q}_{\bar{x}}),$$

such that $\beta^\otimes_{\mathrm{dR}}(s_\alpha \otimes 1) = s_{\alpha,\mathrm{dR},\tilde{x}}$, resp. $\beta^\otimes_{\mathrm{cris}}(s_\alpha \otimes 1) = s_{\alpha,\mathrm{cris},\bar{x}}$[3].The inverse image

$$\beta^{-1}_{\mathrm{dR}}(F^0(\mathcal{Q}_{\tilde{x}})) \subset \Lambda^\vee \otimes_{\mathbb{Z}_p} \mathcal{O}_F$$

of the deRham filtration $F^0(\mathcal{Q}_{\tilde{x}}) = \mathrm{H}^0(\mathcal{Q}_{\tilde{x}}, \Omega^1_{\mathcal{Q}_{\tilde{x}}/\mathcal{O}_F}) \subset \mathrm{H}^1_{\mathrm{dR}}(\mathcal{Q}_{\tilde{x}})$ now gives an $\mathcal{O}_F$-point of the Grassmannian $\mathrm{Gr}(g, \Lambda^\vee)$. This is easily seen to factor through ρ_* to the local model $\mathrm{M}_K(G, \{\mu\})$ and defines the image of λ on $\tilde{x} \in \mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)(\mathcal{O}_F)$.

Next, we use β to construct a versal deformation of the abelian scheme $\mathcal{Q}_{\tilde{x}}$ over the completion $S = \widehat{\mathcal{O}}_{\mathrm{M}_K(G,\{\mu\}),\bar{y}}$ of the local ring of $\mathrm{M}_K(G, \{\mu\})$ at $\bar{y} = \lambda(\bar{x})$. This deformation is equipped with Frobenius invariant crystalline tensors that appropriately extend $\tilde{s}_\alpha$. It is constructed using Zink's theory of displays Zink [2002]. In this, (connected) p-divisible groups over a p-adic ring S are described by "displays", which are systems of Frobenius modules over the ring of Witt vectors $W(S)$. In our case, we give a display over $W(S)$ whose Frobenius semilinear maps are, roughly speaking, valued in the

[3] In the case that K is hyperspecial and $F = F_0$, the existence of such an isomorphism β_{cris} was conjectured by Milne and was shown by Kisin [2010] and Vasiu [2013].

group scheme $\mathcal{G}$. Our construction is inspired by the idea that the local model $\mathrm{M}_K(G, \{\mu\})$ should also appear inside a Witt affine Grassmannian (see Section 2 and also Section 5 for a similar construction).

Finally, using a result of Blasius and Wintenberger and parallel transport for the Gauss-Manin connection, we show that there exists a homomorphism $S \to \widehat{\mathcal{O}}_{\mathcal{S}_{\mathbb{K}}(\mathbb{G},X),\bar{x}}$ which matches $\tilde{x}$ and $\tilde{y}$ and then has to be an isomorphism. Property (c') and then (c) follows.

To deal with the case that $(\mathbb{G}, X)$ is of abelian type, we follow Deligne's strategy Deligne [1979] for constructing canonical models for these types. Deligne relates a connected component of the Shimura variety for $(\mathbb{G}, X)$ to one of the Hodge type $(\mathbb{G}_1, X_1)$ (as in the definition of abelian type above). In this, he uses an action of $\mathbb{G}^{\mathrm{ad}}(\mathbb{Q}) \cap \mathbb{G}^{\mathrm{ad}}(\mathbb{R})^+$ on these varieties. The argument extends after giving a moduli interpretation for this action. Then, $\mathcal{S}_K(\mathbb{G}, X)$ is constructed, by first taking a quotient of a connected component of $\mathcal{S}_{K_1}(\mathbb{G}_1, X_1)$ by a finite group action (which is shown to be free), and then "inducing" from that quotient.

The strategy for these proofs is due to Kisin [2010] who showed the results when K is hyperspecial (*i.e.* $\mathcal{G}$ is reductive). Then $\mathrm{M}_K(G, \{\mu\})$ is the natural smooth model of the homogeneous space X_μ and $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$ is smooth over $\mathcal{O}_E$. (The condition on $\pi_1(G^{\mathrm{der}})$ is not needed then. Also, the case $p = 2$ is treated by Kim and Madapusi Pera [2016].) In this important hyperspecial case, the integral models $\mathcal{S}_K(\mathbb{G}, X)$ can be shown to be "canonical", in the sense of Milne: They can be characterized as the unique, up to isomorphism, regular, formally smooth schemes that satisfy (a), (c'), and an extension property stronger than (b) in which Spec (R) is replaced by any regular, formally smooth scheme over $\mathcal{O}_E$ (see Kisin [2010], Vasiu and Zink [2010]). Hence, they are also independent of the various choices made in their construction. It is unclear if schemes satisfying (a), (b) and (c) as in the conjecture, with $\mathrm{M}_K(G, \{\mu\})$ given a priori, are uniquely determined in the general parahoric case. In Kisin and Pappas [2015], we show this in a few cases. Regardless, we conjecture that the schemes $\mathcal{S}_K(\mathbb{G}, X)$ produced by the above results are independent of the choices made in their construction.

In the hyperspecial case, there is also earlier work of Vasiu [1999] who pursued a different approach (see also Moonen [1998]). Vasiu applied directly an integral comparison homomorphism between p-adic étale cohomology and crystalline cohomology due to Faltings, to understand carefully chosen crystalline tensors of low degree that should be enough to control the group.

Let us note here that, in contrast to all the previous theory, there are few results when K is a subgroup smaller than parahoric, even in PEL cases: When K is the pro-p radical of an Iwahori subgroup, combining the above with results of Oort-Tate or Raynaud on p-torsion group schemes, often leads to well-behaved models (Pappas [1995], Haines and Rapoport [2012]). Also, when the deformation theory is controlled by a formal group (or

"$\mathcal{O}$-module") of dimension 1 (as is in the case of modular curves) the notion of Drinfeld level structure can be used to describe integral models for all level subgroups (Katz and Mazur [1985], Harris and Taylor [2001]).

4 Reductions: Singularities and points

Singularities: Kottwitz-Rapoport stratification and nearby cycles. To fix ideas, we assume that the assumptions of Theorem 3.2 (ii) are satisfied so that the result applies. In particular, $(\mathbb{G}, X)$ is of Hodge type. Set $\mathcal{S} := \mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$, $\mathrm{M} := \mathrm{M}_K(G, \{\mu\})$, $S :=$ Spec $(\mathcal{O}_E)$ and $d := \dim(\mathrm{Sh}_{\mathbb{K}}(\mathbb{G}, X))$.

Since the morphism λ is smooth, Theorem 2.3 implies that special fiber $\mathcal{S} \otimes_{\mathcal{O}_E} k_E$ is reduced, and that we obtain a stratification of $\mathcal{S} \otimes_{\mathcal{O}_E} \bar{\mathbb{F}}_p$ by locally closed smooth strata $\mathcal{S}_{[w]} = \lambda_{\mathbb{K}}^{-1}(S_{[w]})$, for $[w] \in \mathrm{Adm}_{\tilde{K}}(\{\mu\})$. Furthermore:

Theorem 4.1. *For each $[w] \in \mathrm{Adm}_{\tilde{K}}(\{\mu\})$, the Zariski closure $\overline{\mathcal{S}}_{[w]}$ is equal to the union $\bigcup_{[w'] \leq [w]} \mathcal{S}_{[w']}$ and is normal and Cohen-Macaulay.*

Now pick a prime $l \neq p$ and consider the complex of nearby cycles $\mathrm{R}\Psi(\mathcal{S}, \bar{\mathbb{Q}}_l)$ of $\mathcal{S} \to S$ on $\mathcal{S} \otimes_{\mathcal{O}_E} \bar{\mathbb{F}}_p$. This is obtained from the nearby cycles $\mathrm{R}\Psi(\mathrm{M}, \bar{\mathbb{Q}}_l)$ of $\mathrm{M} \to S$ by pulling back along the smooth λ. Set $f = [k_E : \mathbb{F}_p]$ and $q = p^f$. For each $r \geq 1$, the semi-simple trace of Rapoport [1990] defines a function

$$\psi_{\mathbb{K},r} : \mathcal{S}(\mathbb{F}_{q^r}) \to \bar{\mathbb{Q}}_l; \quad \psi_{\mathbb{K},r}(\bar{x}) = \mathrm{Tr}^{\mathrm{ss}}(\mathrm{Frob}_{\bar{x}}, \mathrm{R}\Psi(\mathcal{S}, \bar{\mathbb{Q}}_l)_{\bar{x}}).$$

The smoothness of λ and $\mathcal{G}$, and Lang's lemma, gives that $\psi_{\mathbb{K},r}$ factors through

$$\lambda_{\mathbb{K}}(\mathbb{F}_{q^r}) : \mathcal{S}(\mathbb{F}_{q^r}) \to \mathcal{G}(\mathbb{F}_{q^r}) \backslash \mathrm{M}_K(G, \{\mu\})(\mathbb{F}_{q^r}).$$

Denote by $\mathbb{Q}_{p^n}$ the unramified extension of $\mathbb{Q}_p$ of degree n contained in L; let $\mathbb{Z}_{p^n} \simeq W(\mathbb{F}_{p^n})$ be its integers. Set $K_r = \mathcal{G}(\mathbb{Z}_{q^r})$. If G is quasi-split over $\mathbb{Q}_{q^r}$, then we have a bijection $\mathcal{G}(\mathbb{F}_{q^r}) \backslash \mathrm{M}_K(G, \{\mu\})(\mathbb{F}_{q^r}) \cong K_r \backslash G(\mathbb{Q}_{q^r})/K_r$, and so the map $\lambda_{\mathbb{K}}(\mathbb{F}_{q^r})$ gives $\lambda_{\mathbb{K},r} : \mathcal{S}(\mathbb{F}_{q^r}) \to K_r \backslash G(\mathbb{Q}_{q^r})/K_r$.

By combining results of Pappas and Zhu [2013] on $\mathrm{R}\Psi(\mathrm{M}, \bar{\mathbb{Q}}_l)$ with the above, we obtain:

Theorem 4.2. *a) Suppose that G splits over the finite extension $F/\mathbb{Q}_p$. Then the inertia subgroup $I_{EF} \subset \mathrm{Gal}(\bar{\mathbb{Q}}_p/EF)$ acts unipotently on $\mathrm{R}\Psi(\mathcal{S}, \bar{\mathbb{Q}}_l)$.*

b) Suppose that G is quasi-split over $\mathbb{Q}_{q^r}$. Then the semi-simple trace of Frobenius $\psi_{r,\mathbb{K}} : \mathcal{S}(\mathbb{F}_{q^r}) \to \bar{\mathbb{Q}}_l$ is a composition

$$\mathcal{S}(\mathbb{F}_{q^r}) \xrightarrow{\lambda_{\mathbb{K},r}} K_r \backslash G(\mathbb{Q}_{q^r})/K_r \xrightarrow{z_{\mu,r}} \bar{\mathbb{Q}}_l$$

where $z_{\mu,r}$ is in the center of the Hecke algebra $\bar{\mathbb{Q}}_l[K_r\backslash G(\mathbb{Q}_{q^r})/K_r]$ with multiplication by convolution.

c) Suppose that G is split over $\mathbb{Q}_{q^r}$. Then in (b) above, we can take $z_{\mu,r}$ such that $q^{-d/2}z_{\mu,r}$ is a Bernstein function for μ.

Part (c) has been conjectured by Kottwitz, see Rapoport [2005, (10.3)] and also work of Haines [2014] for more details and an extension of the conjecture. Let us mention that the study of $\mathrm{R}\Psi(\mathrm{M},\bar{\mathbb{Q}}_l)$ in Pappas and Zhu [2013] (as also the proof of Theorem 2.3), uses techniques from the theory of the geometric Langlands correspondence. The "Hecke central" statement was first shown by Gaitsgory for split groups over function fields Gaitsgory [2001], and by Haines and Ngô for unramified unitary and symplectic groups Haines and Ngô [2002]. Gaitsgory's result was, in turn, inspired by Kottwitz's conjecture.

Points modulo p. Under certain assumptions, Langlands and Rapoport gave a conjectural description of the set of $\bar{\mathbb{F}}_p$-points of a model of a Shimura variety together with its actions by Frobenius and Hecke operators Langlands and Rapoport [1987]. The idea, very roughly, is as follows (see also Langlands [1976], Kottwitz [1992]). Since, as suggested by Deligne, most Shimura varieties are supposed to be moduli spaces of certain motives, we should be describing this set via representations of the fundamental groupoid attached to the Tannakian category of motives over $\bar{\mathbb{F}}_p$. A groupoid $\mathfrak{Q}$ which should be this fundamental groupoid (and almost is, assuming the Tate conjecture and other standard conjectures) can be constructed explicitly. Then "$\mathbb{G}$-pseudo-motives" are given by groupoid homomorphisms $\varphi:\mathfrak{Q}\to\mathbb{G}$. The choice of the domain X imposes restrictions, and we should be considering only φ which are "admissible". Then, we also give p and prime-to-p level structures on these φ, as we do for abelian varieties.

The Langlands–Rapoport conjecture was corrected, modified and extended along the way (Reimann [1997] and Kisin [2017]). The conjecture makes better sense when it refers to a specific integral model of the Shimura variety $\mathrm{Sh}_K(\mathbb{G},X)$. When K is hyperspecial, p odd, and $(\mathbb{G},X)$ of abelian type, an extended version of the conjecture was essentially proven (with a caveat, see below) by Kisin [2017], for the canonical integral models constructed in Kisin [2010]. In fact, the conjecture also makes sense when K is parahoric, for the integral models of Kisin and Pappas [2015]. More precisely, let us suppose that the assumptions of Theorem 3.2 are satisfied and that $\mathscr{S}_K(\mathbb{G},X)$ is provided by the construction in the proof.

Conjecture 4.3. *There is a $\langle\Phi_E\rangle\times\mathbb{G}(\mathbb{A}_f^p)$-equivariant bijection*

$$(LR)\qquad \mathscr{S}_K(\mathbb{G},X)(\bar{\mathbb{F}}_p)\xrightarrow{\sim}\bigsqcup_{[\varphi]}\varprojlim_{K^p} I_\varphi(\mathbb{Q})\backslash(X_p(\varphi)\times(X^p(\varphi)/K^p))$$

where Φ_E is the Frobenius over k_E and the rest of the notations and set-up follow Rapoport [2005, §9], Kisin [2017, (3.3)] (taking into account the remark (3.3.9) there).

In particular: The disjoint union is over a set of equivalence classes $[\varphi]$ of admissible $\varphi : \mathfrak{Q} \to \mathbb{G}$, where one uses Kisin's definition of "admissible" for non simply connected derived group. (The classes $[\varphi]$ often correspond to isogeny classes of abelian varieties with additional structures.) We have $I_\varphi = \mathrm{Aut}(\varphi)$, an algebraic group over $\mathbb{Q}$. The set $X^p(\varphi)$ is a right $\mathbb{G}(\mathbb{A}_f^p)$-torsor and the set $X_p(\varphi)$ can be identified with the affine Deligne-Lusztig set $X_K(\{\mu^{-1}\}, b)$ for $b \in G(L)$ obtained from φ. The group $I_\varphi(\mathbb{Q})$ acts on $X^p(\varphi)$ on the left, and there is an injection $I_\varphi(\mathbb{Q}) \to J_b(\mathbb{Q}_p)$ which also produces a left action on $X_p(\varphi) = X_K(\{\mu^{-1}\}, b)$; the quotient is by the diagonal action.

Theorem 4.4. *(Kisin [ibid.]) Assume that K is hyperspecial, p odd, and $(\mathbb{G}, X)$ of abelian type. Then there is a bijection as in (LR) respecting the action of $\langle \Phi_E \rangle \times \mathbb{G}(\mathbb{A}_f^p)$ on both sides, but with the action of $I_\varphi(\mathbb{Q})$ on $X_p(\varphi) \times (X^p(\varphi)/K^p)$ obtained from the natural diagonal action above by conjugating by a (possibly trivial) element $\tau(\varphi) \in I_\varphi^{\mathrm{ad}}(\mathbb{A}_f)$.*

Due to lack of space, we will omit an account of any of the beautiful and subtle arguments in Kisin's proof, or in related earlier work by Kottwitz [1992] and others. We hope that the interested reader will consult the original papers. Let us just mention here that Zhou [2017] has recently made some progress towards the proof of Conjecture 4.3.

5 Local theory: Formal schemes

We now return to the local set-up as in Section 1. Let $(G, \{\mu\})$ be a local Shimura pair and $[b] \in B(G)$ a σ-conjugacy class. Fix a parahoric subgroup $K \subset G(\mathbb{Q}_p)$ and assume that the local model $\mathrm{M}_K(G, \{\mu^{-1}\})$ is defined.

Conjecture 5.1. *There exists a formal scheme $\mathfrak{X}_K(\{\mu\}, b)$ over* $\mathrm{Spf}(\mathcal{O}_E)$ *with $J_b(\mathbb{Q}_p)$-action, which is locally formally of finite type, and is such that:*

a) There is a $\langle \Phi_E \rangle \times J_b(\mathbb{Q}_p)$-equivariant bijection between $\mathfrak{X}_K(\{\mu\}, b)(\bar{\mathbb{F}}_p)$ and the affine Deligne-Lusztig set $X_K(\{\mu\}, b)$.

b) For any $v \in X_K(\{\mu\}, b)$, there is $w \in \mathrm{M}_K(G, \{\mu^{-1}\})(\bar{\mathbb{F}}_p)$ such that the completion of $\mathfrak{X}_K(\{\mu\}, b)$ at the point corresponding to v via (a) is isomorphic to the completion of $\mathrm{M}_K(G, \{\mu^{-1}\})$ at w.

Assuming $\mathfrak{X}_K(\{\mu\}, b)$ exists, we let $\mathcal{X}_K(\{\mu\}, b)$ be its underlying reduced scheme which is locally of finite type over Spec (k_E). (Then $\mathcal{X}_K(\{\mu\}, b)$ could be called an *affine Deligne-Lusztig "variety"*.) Its "perfection" $\mathcal{X}_K(\{\mu\}, b)^{\mathrm{perf}}$, has been constructed by Zhu [2017], using a Witt vector affine flag variety. The rigid fiber $\mathfrak{X}_K(\{\mu\}, b)^{\mathrm{rig}}$ over E would be the *local Shimura variety* for $(G, \{\mu\}, [b])$ and level K whose existence is expected by Rapoport and Viehmann [2014].

Let us take $G = \mathrm{GL}_n$, $\mu_d(z) = \mathrm{diag}(z, \dots, z, 1, \dots, 1)^{-1}$, with d copies of z, and $K = \mathrm{GL}_n(\mathbb{Z}_p)$. For simplicity, set $W = W(\bar{\mathbb{F}}_p)$. The formal schemes $\mathfrak{X}_{\mathrm{GL}_n} :=$

$\mathfrak{X}_K(\{\mu_d\}, b)$ were constructed by Rapoport and Zink [1996]: Fix a p-divisible group H_0 of height n and dimension d over $\bar{\mathbb{F}}_p$ with Frobenius $F = b\cdot\sigma$ on the rational Dieudonné module. Then the formal scheme $\mathfrak{X}_{\mathrm{GL}_n}\hat{\otimes}_{\mathbb{Z}_p} W$ represents (Rapoport and Zink [ibid., Theorem 2.16]) the functor which sends a W-algebra R with p nilpotent on R, to the set of isomorphism classes of pairs (H, τ), where: H is a p-divisible group over R, and $\tau : H_0 \otimes_k R/pR \dashrightarrow H \otimes_R R/pR$ is a quasi-isogeny. The formal scheme $\mathfrak{X}_{\mathrm{GL}_n}$ over $\mathrm{Spf}(\mathbb{Z}_p)$ is obtained from this by descent. This construction of $\mathfrak{X}_K(\{\mu\}, b)$ generalizes to $(G, \{\mu\})$ that are of "EL" or "PEL" type, for many parahoric K; these types are defined similarly to (and often arise from) global Shimura data $(\mathbb{G}, X)$ of PEL type.

The models $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$ of Conjecture 3.1 and the formal schemes $\mathfrak{X}_K(\{\mu\}, b)$ above, should be intertwined via the bijection (LR) as follows:

Suppose that $(\mathbb{G}, X)$ is a Shimura datum which produces the local Shimura pair $(G, \{\mu^{-1}\})$. Take $\bar{x} \in \mathcal{S}_K(\mathbb{G}, X)(\bar{\mathbb{F}}_p)$ with corresponding $b \in G(L)$. Then, there should be a morphism of $\mathrm{Spf}(\mathcal{O}_E)$–formal schemes

$$i_{\bar{x}} : \mathfrak{X}_K(\{\mu\}, b) \times \mathbb{G}(\mathbb{A}_f^p)/K^p \to \widehat{\mathcal{S}}_{\mathbb{K}}(\mathbb{G}, X) := \varprojlim_n \mathcal{S}_{\mathbb{K}}(\mathbb{G}, X) \otimes_{\mathcal{O}_E} \mathcal{O}_E/(p^n),$$

which, on $\bar{\mathbb{F}}_p$-points is given by (LR) and surjects on the "isogeny class" $[\varphi_0]$ of $\bar{x}$, and induces isomorphisms on the formal completions at closed points.

An interesting special case is when the σ-conjugacy class $[b]$ is basic (Kottwitz [1997]). Then the image Z of $i_{\bar{x}}$ should be closed in the special fiber of $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$ and $i_{\bar{x}}$ should be giving a "p-adic uniformization" (*cf.* Rapoport and Zink [1996])

$$(U) \qquad I_{\varphi_0}(\mathbb{Q})\backslash\mathfrak{X}_K(\{\mu\}, b) \times (\mathbb{G}(\mathbb{A}_f^p)/K^p) \xrightarrow{\sim} \widehat{\mathcal{S}}_{\mathbb{K}}(\mathbb{G}, X)_{/Z},$$

of the completion of $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$ along Z.

Assume that $(\mathbb{G}, X)$ is of Hodge type and that the rest of the assumptions of Theorem 3.2 (ii) are also satisfied. Then we can hope to construct $\mathfrak{X}_K(\{\mu\}, b)$ using the model $\mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)$ as follows: First consider the fiber product

$$\begin{array}{ccc} \mathfrak{F} & \to & (\mathfrak{X}_{\mathrm{GSp}} \times 1)\hat{\otimes}_{\mathbb{Z}_p}\mathcal{O}_E \\ \downarrow & & \downarrow i_{\bar{x}} \\ \widehat{\mathcal{S}}_{\mathbb{K}}(\mathbb{G}, X) & \xrightarrow{\iota} & \widehat{\mathcal{S}}_{\mathbb{K}'}(\mathrm{GSp}(\mathbb{V}), S^{\pm})\hat{\otimes}_{\mathbb{Z}_p}\mathcal{O}_E, \end{array}$$

where ι is an appropriate Hodge embedding as in Kisin [2010], or Kisin and Pappas [2015], and $\mathfrak{X}_{\mathrm{GSp}}$ is the Rapoport-Zink formal scheme for the symplectic PEL type. Then, provided we have the set map $X_K(\{\mu\}, b) \to \mathcal{S}_{\mathbb{K}}(\mathbb{G}, X)(\bar{\mathbb{F}}_p)$ sending $1\cdot\tilde{K}$ to $\bar{x}$ (as predicted by the (LR) conjecture), we can define $\mathfrak{X}_K(\{\mu\}, b)$ to be the formal completion of $\mathfrak{F}$ along the (closed) subset given by $X_K(\{\mu\}, b)$. The existence of the "uniformization" morphism $i_{\bar{x}}$ follows immediately. Howard and the author applied this idea to show:

Theorem 5.2. *(Howard and Pappas [2017]) Assume $p \neq 2$, $(\mathbb{G}, X)$ is of Hodge type and K is hyperspecial. Choose a Hodge embedding ρ, a lattice Λ, and tensors $\{s_\alpha\}_\alpha$, as in the proof of Theorem 3.2 (ii). Suppose $\bar{x} \in \mathcal{S}_{\mathbb{K}}(G, X)(\bar{\mathbb{F}}_p)$ and b are as above. Then, $\mathfrak{X}_K(\{\mu\}, b)$ satisfying Conjecture 5.1 can be defined as above, and only depends, up to isomorphism, on $(G, \{\mu\}, b, K)$ and $\rho_{\mathbb{Z}_p} : \mathcal{G} \to \mathrm{GL}(\Lambda)$.*

For basic $[b]$, the uniformization (U) also follows. In the above result, the formal scheme $\mathfrak{X}_K(\{\mu\}, b)\hat{\otimes}_{\mathcal{O}_E} W$ represents a functor on the category of formally finitely generated adic W-algebras (A, I) which are formally smooth over $W/(p^m)$, for some $m \geq 1$ as follows. Set H_0 for the p-divisible group $\mathcal{A}_{\bar{x}}[p^\infty]$. We send (A, I) to the set of isomorphism classes of triples $(H, \tau, \{t_\alpha\}_\alpha)$, where H is a p-divisible group over A, $\tau : H_0 \otimes_k A/(p, I) \dashrightarrow H \otimes_A A/(p, I)$ is a quasi-isogeny, and t_α are (Frobenius invariant) sections of the Dieudonné crystal $\mathbb{D}(H)^\otimes$. We require that the pull-backs $\tau^*(t_\alpha)$ agree in $\mathbb{D}(H_0 \otimes_k A/(p, I))^\otimes[1/p]$ with the constant sections given by $s_{\alpha,\mathrm{crys},\bar{x}}$, and also some more properties listed in Howard and Pappas [ibid., (2.3.3), (2.3.6)]. For K hyperspecial, a different, local, construction of the formal scheme $\mathfrak{X}_K(\{\mu\}, b)$ (still only for "local Hodge types") first appeared in work of Kim [2013].

In some cases, we can also give $\mathfrak{X}_K(\{\mu\}, b)$ directly, by using a group theoretic version of Zink's displays. Assume that K is hyperspecial, so $\mathcal{G}$ is reductive over $\mathbb{Z}_p$. Set $\mathcal{G}_W = \mathcal{G} \otimes_{\mathbb{Z}_p} W$ and pick $\mu : \mathbb{G}_{\mathrm{m}W} \to \mathcal{G}_W$ in our conjugacy class. Set $L^+\mathcal{G}$, resp. $L\mathcal{G}$, for the "positive Witt loop" group scheme, resp. "Witt loop" ind-group scheme, representing $R \mapsto \mathcal{G}(W(R))$, resp. $R \mapsto \mathcal{G}(W(R)[1/p])$. Let $\mathcal{H}_\mu$ be the subgroup scheme of $L^+\mathcal{G}_W$ with R-points given by $g \in \mathcal{G}(W(R))$ whose projection $g_0 \in \mathcal{G}(R)$ lands in the R-points of the parabolic subgroup scheme $\mathcal{P}_\mu \subset \mathcal{G}_W$ associated to μ. We can define a homomorphism $\Phi_{G,\mu} : \mathcal{H}_\mu \to L^+\mathcal{G}_W$ such that $\Phi_{G,\mu}(h) = F \cdot (\mu(p) \cdot h \cdot \mu(p)^{-1})$ in $\mathcal{G}(W(R)[1/p])$, with F given by the Frobenius $W(R) \to W(R)$ (Bueltel and Pappas [2017]). Let R be a p-nilpotent W-algebra. Set

$$\mathfrak{X}'_{\mathcal{G}}(R) = \{(U, g) \in L^+\mathcal{G}(R) \times L\mathcal{G}(R) \mid g^{-1}bF(g) \overset{(*)}{=} U\mu^\sigma(p)\}/\mathcal{H}_\mu(R),$$

where $(*)$ is taken in $L\mathcal{G}(R)$ and the quotient is for the action given by

$$(U, g) \cdot h = (h^{-1}U\Phi_{G,\mu}(h), gh).$$

Denote by $\mathfrak{X}_{\mathcal{G}}$ the étale sheaf associated to $R \mapsto \mathfrak{X}'_{\mathcal{G}}(R)$.

If R is perfect, then F is an isomorphism and $W(R)$ is p-torsion free, so a pair (U, g) with $g^{-1}bF(g) = U\mu^\sigma(p)$ is determined by g. Then

$$\mathfrak{X}'_{\mathcal{G}}(R) = \{g \in L\mathcal{G}(R)/L^+\mathcal{G}(R) \mid g^{-1}bF(g) \in L^+\mathcal{G}(R)\mu^\sigma(p)L^+\mathcal{G}(R)\}.$$

In particular, $\mathfrak{X}_{\mathcal{G}}(\bar{\mathbb{F}}_p) = \mathfrak{X}'_{\mathcal{G}}(\bar{\mathbb{F}}_p) \cong X_K(\{\mu\}, b)$ and the perfection of $\mathfrak{X}_{\mathcal{G}}$ agrees with the space $\mathcal{X}_K(\{\mu\}, b)^{\mathrm{perf}}$ of Zhu [2017]. If $\mathfrak{X}_{\mathcal{G}}$ is representable by a formal scheme, then this satisfies Conjecture 5.1 and gives $\mathfrak{X}_K(\{\mu\}, b)\hat{\otimes}_{\mathcal{O}_E} W$.

Theorem 5.3. *(Bueltel and Pappas [2017]) Let $\rho_{\mathbb{Z}_p} : \mathcal{G} \hookrightarrow \mathrm{GL}_n$ be a closed group scheme embedding. Suppose $\rho_{\mathbb{Z}_p} \circ \mu$ is minuscule and $\rho_{\mathbb{Z}_p}(b)$ has no zero slopes. Then the restriction of $\mathfrak{X}_{\mathcal{G}}$ to Noetherian p-nilpotent W-algebras is represented by a formal scheme which is formally smooth and locally formally of finite type over* $\mathrm{Spf}(W)$.

The proof also gives that, if $\mathfrak{X}_K(\{\mu\}, b)$ is defined as in Theorem 5.2 and $\rho_{\mathbb{Z}_p}(b)$ has no zero slopes, then $\mathfrak{X}_{\mathcal{G}}(R) = \mathfrak{X}_K(\{\mu\}, b)(R)$, for R Noetherian. Hence, then $\mathfrak{X}_K(\{\mu\}, b)$ of Theorem 5.2 is independent of $\rho_{\mathbb{Z}_p}$.

References

Arnaud Beauville and Yves Laszlo (1994). "Conformal blocks and generalized theta functions". *Comm. Math. Phys.* 164.2, pp. 385–419. MR: 1289330 (cit. on p. 400).

A Beilinson and V Drinfeld (1996). "Quantization of Hitchin's Hamiltonians and Hecke eigen–sheaves". *preprint* (cit. on p. 400).

Bhargav Bhatt and Peter Scholze (2017). "Projectivity of the Witt vector affine Grassmannian". *Invent. Math.* 209.2, pp. 329–423. MR: 3674218 (cit. on p. 402).

M. V. Borovoĭ (1987). "Conjugation of Shimura varieties". In: *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*. Amer. Math. Soc., Providence, RI, pp. 783–790. MR: 934281 (cit. on p. 396).

F. Bruhat and J. Tits (1972). "Groupes réductifs sur un corps local". *Inst. Hautes Études Sci. Publ. Math.* 41, pp. 5–251. MR: 0327923 (cit. on p. 398).

– (1984). "Groupes réductifs sur un corps local. II. Schémas en groupes. Existence d'une donnée radicielle valuée". *Inst. Hautes Études Sci. Publ. Math.* 60, pp. 197–376. MR: 756316 (cit. on pp. 398, 400).

O. Bueltel and Georgios Pappas (Feb. 2017). "(G, μ)-displays and Rapoport-Zink spaces". arXiv: 1702.00291 (cit. on pp. 413, 414).

Pierre Deligne (1971). "Travaux de Shimura", 123–165. Lecture Notes in Math., Vol. 244. MR: 0498581 (cit. on p. 395).

– (1979). "Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques". In: *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*. Proc. Sympos. Pure Math., XXXIII. Amer. Math. Soc., Providence, R.I., pp. 247–289. MR: 546620 (cit. on pp. 395, 408).

Pierre Deligne, James S. Milne, Arthur Ogus, and Kuang-yen Shih (1982). *Hodge cycles, motives, and Shimura varieties*. Vol. 900. Lecture Notes in Mathematics. Springer-Verlag, Berlin-New York, pp. ii+414. MR: 654325 (cit. on p. 395).

Pierre Deligne and Georgios Pappas (1994). "Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant". *Compositio Math.* 90.1, pp. 59–79. MR: 1266495 (cit. on p. 404).

D. Gaitsgory (2001). "Construction of central elements in the affine Hecke algebra via nearby cycles". *Invent. Math.* 144.2, pp. 253–280. MR: 1826370 (cit. on p. 410).

Ulrich Görtz (2003). "On the flatness of local models for the symplectic group". *Adv. Math.* 176.1, pp. 89–115. MR: 1978342 (cit. on p. 404).

Thomas J. Haines (2014). "The stable Bernstein center and test functions for Shimura varieties". In: *Automorphic forms and Galois representations. Vol. 2*. Vol. 415. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 118–186. MR: 3444233 (cit. on p. 410).

Thomas J. Haines and B. C. Ngô (2002). "Nearby cycles for local models of some Shimura varieties". *Compositio Math.* 133.2, pp. 117–150. MR: 1923579 (cit. on p. 410).

Thomas J. Haines and Michael Rapoport (2008). "On parahoric subgroups". arXiv: 0804.3788 (cit. on pp. 398, 399).

– (2012). "Shimura varieties with Γ_1(p)-level via Hecke algebra isomorphisms: the Drinfeld case". In: *Annales scientifiques de l'École Normale Supérieure*. Vol. 45. 4. Société mathématique de France, pp. 719–785 (cit. on p. 408).

Michael Harris and Richard Taylor (2001). *The geometry and cohomology of some simple Shimura varieties*. Vol. 151. Annals of Mathematics Studies. With an appendix by Vladimir G. Berkovich. Princeton University Press, Princeton, NJ, pp. viii+276. MR: 1876802 (cit. on pp. 396, 409).

Benjamin Howard and Georgios Pappas (2017). "Rapoport-Zink spaces for spinor groups". *Compos. Math.* 153.5, pp. 1050–1118. MR: 3705249 (cit. on p. 413).

A. J. de Jong (1993). "The moduli spaces of principally polarized abelian varieties with $\Gamma_0(p)$-level structure". *J. Algebraic Geom.* 2.4, pp. 667–688. MR: 1227472 (cit. on p. 404).

Nicholas M. Katz and Barry Mazur (1985). *Arithmetic moduli of elliptic curves*. Vol. 108. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xiv+514. MR: 772569 (cit. on p. 409).

Wansu Kim (Aug. 2013). "Rapoport-Zink spaces of Hodge type". arXiv: 1308.5537 (cit. on p. 413).

Wansu Kim and Keerthi Madapusi Pera (2016). "2-adic integral canonical models". *Forum Math. Sigma* 4, e28, 34. MR: 3569319 (cit. on p. 408).

Mark Kisin (2009). "Moduli of finite flat group schemes, and modularity". *Ann. of Math. (2)* 170.3, pp. 1085–1180. MR: 2600871 (cit. on pp. 397, 406, 407).

– (2010). "Integral models for Shimura varieties of abelian type". *J. Amer. Math. Soc.* 23.4, pp. 967–1012. MR: 2669706 (cit. on pp. 397, 405–408, 410, 412).

Mark Kisin (2017). "mod p points on Shimura varieties of abelian type". *J. Amer. Math. Soc.* 30.3, pp. 819–914. MR: 3630089 (cit. on pp. 410, 411).

Mark Kisin and Georgios Pappas (Dec. 2015). "Integral models of Shimura varieties with parahoric level structure". To appear in Publ. Math. IHES. arXiv: 1512.01149 (cit. on pp. 397, 405, 408, 410, 412).

Robert E. Kottwitz (1990). "Shimura varieties and λ-adic representations". In: *Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988)*. Vol. 10. Perspect. Math. Academic Press, Boston, MA, pp. 161–209. MR: 1044820 (cit. on p. 396).

– (1992). "Points on some Shimura varieties over finite fields". *J. Amer. Math. Soc.* 5.2, pp. 373–444. MR: 1124982 (cit. on pp. 396, 402, 410, 411).

– (1997). "Isocrystals with additional structure. II". *Compositio Math.* 109.3, pp. 255–339. MR: 1485921 (cit. on pp. 398, 412).

R. P. Langlands (1976). "Some contemporary problems with origins in the Jugendtraum", pp. 401–418. MR: 0437500 (cit. on p. 410).

– (1979). "Sur la mauvaise réduction d'une variété de Shimura". In: *Journées de Géométrie Algébrique de Rennes. (Rennes, 1978), Vol. III*. Vol. 65. Astérisque. Soc. Math. France, Paris, pp. 125–154. MR: 563475 (cit. on p. 396).

R. P. Langlands and Michael Rapoport (1987). "Shimuravarietäten und Gerben". *J. Reine Angew. Math.* 378, pp. 113–220. MR: 895287 (cit. on p. 410).

Robert P. Langlands and Dinakar Ramakrishnan, eds. (1992). *The zeta functions of Picard modular surfaces*. Université de Montréal, Centre de Recherches Mathématiques, Montreal, QC, pp. xiv+492. MR: 1155223 (cit. on p. 396).

Brandon Levin (2016). "Local models for Weil-restricted groups". *Compos. Math.* 152.12, pp. 2563–2601. MR: 3594288 (cit. on p. 401).

J. S. Milne (1990). "Canonical models of (mixed) Shimura varieties and automorphic vector bundles". In: *Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988)*. Vol. 10. Perspect. Math. Academic Press, Boston, MA, pp. 283–414. MR: 1044823 (cit. on pp. 396, 403).

Ben Moonen (1998). "Models of Shimura varieties in mixed characteristics". In: *Galois representations in arithmetic algebraic geometry (Durham, 1996)*. Vol. 254. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 267–350. MR: 1696489 (cit. on p. 408).

Georgios Pappas (1995). "Arithmetic models for Hilbert modular varieties". *Compositio Math.* 98.1, pp. 43–76. MR: 1353285 (cit. on p. 408).

– (2000). "On the arithmetic moduli schemes of PEL Shimura varieties". *J. Algebraic Geom.* 9.3, pp. 577–605. MR: 1752014 (cit. on p. 401).

Georgios Pappas and Michael Rapoport (2008). "Twisted loop groups and their affine flag varieties". *Adv. Math.* 219.1. With an appendix by T. Haines and Rapoport, pp. 118–198. MR: 2435422 (cit. on pp. 399, 401).

Georgios Pappas, Michael Rapoport, and Brian Smithling (2013). "Local models of Shimura varieties, I. Geometry and combinatorics". In: *Handbook of moduli. Vol. III*. Vol. 26. Adv. Lect. Math. (ALM). Int. Press, Somerville, MA, pp. 135–217. MR: 3135437 (cit. on pp. 399, 401).

Georgios Pappas and X. Zhu (2013). "Local models of Shimura varieties and a conjecture of Kottwitz". *Invent. Math.* 194.1, pp. 147–254. MR: 3103258 (cit. on pp. 400, 401, 404, 405, 409, 410).

Michael Rapoport (1990). "On the bad reduction of Shimura varieties". In: *Automorphic forms, Shimura varieties, and L-functions, Vol. II (Ann Arbor, MI, 1988)*. Vol. 11. Perspect. Math. Academic Press, Boston, MA, pp. 253–321. MR: 1044832 (cit. on pp. 396, 409).

– (2005). "A guide to the reduction modulo p of Shimura varieties". *Astérisque* 298, pp. 271–318 (cit. on pp. 399, 410).

Michael Rapoport and Eva Viehmann (2014). "Towards a theory of local Shimura varieties". *Münster J. Math.* 7.1, pp. 273–326. MR: 3271247 (cit. on pp. 398, 411).

Michael Rapoport and Th. Zink (1982). "Über die lokale Zetafunktion von Shimuravarietäten. Monodromiefiltration und verschwindende Zyklen in ungleicher Charakteristik". *Invent. Math.* 68.1, pp. 21–101. MR: 666636 (cit. on p. 396).

– (1996). *Period spaces for p-divisible groups*. Vol. 141. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xxii+324. MR: 1393439 (cit. on pp. 397, 401, 402, 404, 412).

Harry Reimann (1997). *The semi-simple zeta function of quaternionic Shimura varieties*. Vol. 1657. Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. viii+143. MR: 1470457 (cit. on p. 410).

Peter Scholze (2013). "The Langlands-Kottwitz method and deformation spaces of p-divisible groups". *J. Amer. Math. Soc.* 26.1, pp. 227–259. MR: 2983011 (cit. on p. 396).

Goro Shimura (1964). "On the field of definition for a field of automorphic functions". *Ann. of Math. (2)* 80, pp. 160–189. MR: 0167466 (cit. on p. 395).

– (1970). "On canonical models of arithmetic quotients of bounded symmetric domains. II". *Ann. of Math. (2)* 92, pp. 528–549. MR: 0292758 (cit. on p. 395).

J. Tits (1979). "Reductive groups over local fields". In: *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1*. Proc. Sympos. Pure Math., XXXIII. Amer. Math. Soc., Providence, R.I., pp. 29–69. MR: 546588 (cit. on p. 398).

Adrian Vasiu (1999). "Integral canonical models of Shimura varieties of preabelian type". *Asian J. Math.* 3.2, pp. 401–518. MR: 1796512 (cit. on pp. 397, 408).

Adrian Vasiu (2013). “A motivic conjecture of Milne”. *J. Reine Angew. Math.* 685, pp. 181–247. MR: 3181569 (cit. on p. 407).

Adrian Vasiu and Thomas Zink (2010). “Purity results for p-divisible groups and abelian schemes over regular bases of mixed characteristic”. *Doc. Math.* 15, pp. 571–599. MR: 2679067 (cit. on p. 408).

Rong Zhou (July 2017). “Mod-p isogeny classes on Shimura varieties with parahoric level structure”. arXiv: 1707.09685 (cit. on p. 411).

Xinwen Zhu (2014). “On the coherence conjecture of Pappas and Rapoport”. *Ann. of Math. (2)* 180.1, pp. 1–85. MR: 3194811 (cit. on p. 401).

– (2017). “Affine Grassmannians and the geometric Satake in mixed characteristic”. *Ann. of Math. (2)* 185.2, pp. 403–492. MR: 3612002 (cit. on pp. 402, 411, 413).

Thomas Zink (2002). “The display of a formal p-divisible group”. *Astérisque* 278. Cohomologies p-adiques et applications arithmétiques, I, pp. 127–248. MR: 1922825 (cit. on p. 407).

Received 2017-11-29.

Georgios Pappas
pappas@math.msu.edu
Dept. of Mthematics
Michigan State University
E. Lansing, MI 48824
USA

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (417–432)

HEURISTICS FOR THE ARITHMETIC OF ELLIPTIC CURVES

Bjorn Poonen

Abstract

This is an introduction to a probabilistic model for the arithmetic of elliptic curves, a model developed in a series of articles of the author with Bhargava, Kane, Lenstra, Park, Rains, Voight, and Wood. We discuss the theoretical evidence for the model, and we make predictions about elliptic curves based on corresponding theorems proved about the model. In particular, the model suggests that all but finitely many elliptic curves over $\mathbb{Q}$ have rank ≤ 21, which would imply that the rank is uniformly bounded.

1 Introduction

Let E be an elliptic curve over $\mathbb{Q}$ (see Silverman [2009] for basic definitions). Let $E(\mathbb{Q})$ be the set of rational points on E. The group law on E gives $E(\mathbb{Q})$ the structure of an abelian group, and Mordell [1922] proved that $E(\mathbb{Q})$ is finitely generated; let $\operatorname{rk} E(\mathbb{Q})$ denote its rank. The present survey article, based primarily on articles Poonen and Rains [2012], Bhargava, Kane, Lenstra, Poonen, and Rains [2015], and Park, Poonen, Voight, and Wood [2016], is concerned with the following question:

Question 1.1. Is $\operatorname{rk} E(\mathbb{Q})$ bounded as E varies over all elliptic curves over $\mathbb{Q}$?

Question 1.1 was implicitly asked by Poincaré [1901, p. 173] in 1901, even before $E(\mathbb{Q})$ was known to be finitely generated! Since then, many authors have put forth guesses, and the folklore expectation has flip-flopped at least once; see Poincaré [1950, p. 495, end of footnote (^{3})], Honda [1960, p. 98], Cassels [1966, p. 257], Tate [1974, p. 194], Mestre [1982], Mestre [1986, II.1.1 and II.1.2], Brumer [1992, Section 1], Ulmer [2002, Conjecture 10.5], and Farmer, Gonek, and Hughes [2007, (5.20)], or see Park, Poonen, Voight, and Wood [2016, Section 3.1] for a summary.

The writing of this article was supported in part by National Science Foundation grant DMS-1601946 and Simons Foundation grants #402472 (to Bjorn Poonen) and #550033.
MSC2010: primary 11G05; secondary 11G40, 14G25, 14H52, 14K15.
Keywords: Elliptic curve, rank, Selmer group, Shafarevich–Tate group, abelian variety.

The present survey describes a probabilistic model for the arithmetic of elliptic curves, and presents theorems about the model that suggest that $\operatorname{rk} E(\mathbb{Q}) \le 21$ for all but finitely many elliptic curves E, and hence that $\operatorname{rk} E(\mathbb{Q})$ is bounded. Ours is not the first heuristic for boundedness: there is one by Rubin and Silverberg [2000, Remarks 5.1 and 5.2], for a family of quadratic twists and another by Granville, discussed in Watkins, Donnelly, Elkies, Fisher, Granville, and Rogers [2014, Section 11] and developed further in Watkins [2015]. Interestingly, the latter also suggests a bound of 21.

Modeling ranks directly is challenging because there are few theorems about the distribution of ranks. Also, although there exists extensive computational data that suggests answers to some questions (e.g., Balakrishnan, Ho, Kaplan, Spicer, Stein, and Weigandt [2016]), it seems that far more data would be needed to suggest answers to others. Therefore, instead of modeling ranks in isolation, we model ranks, Selmer groups, and Shafarevich–Tate groups simultaneously, so that we can calibrate and corroborate the model using a diverse collection of known results.

2 The arithmetic of elliptic curves

2.1 Counting elliptic curves by height. Every elliptic curve E over $\mathbb{Q}$ is isomorphic to the projective closure of a unique curve $y^2 = x^3 + Ax + B$ in which A and B are integers with $4A^3 + 27B^2 \neq 0$ (the smoothness condition) such that there is no prime p such that $p^4|A$ and $p^6|B$. Let $\mathscr{E}$ be the set of elliptic curves of this form, so $\mathscr{E}$ contains one curve in each isomorphism class. Define the height of $E \in \mathscr{E}$ by

$$\operatorname{ht} E := \max(|4A^3|, |27B^2|).$$

(This definition is specific to the ground field $\mathbb{Q}$, but it has analogues over other number fields.) Define

$$\mathscr{E}_{\le H} := \{E \in \mathscr{E} : \operatorname{ht} E \le H\}.$$

Ignoring constant factors, we have about $H^{1/3}$ integers A with $|4A^3| \le H$, and $H^{1/2}$ integers B with $|27B^2| \le H$. A positive fraction of such pairs (A, B) satisfy the smoothness condition and divisibility conditions, so one should expect $\#\mathscr{E}_{\le H}$ to be about $H^{1/3}H^{1/2} = H^{5/6}$. In fact, an elementary sieve argument [Brumer 1992, Lemma 4.3] proves the following:

Proposition 2.1. *We have*

$$\#\mathscr{E}_{\le H} = \left(2^{4/3}3^{-3/2}\zeta(10)^{-1} + o(1)\right) H^{5/6}$$

as $H \to \infty$.

Define the density of a subset $S \subseteq \mathscr{E}$ as

$$\lim_{H \to \infty} \frac{\#(S \cap \mathscr{E}_{\leq H})}{\#\mathscr{E}_{\leq H}},$$

if the limit exists. For example, it is a theorem that 100% of elliptic curves E over $\mathbb{Q}$ have no nontrivial rational torsion points; this statement is to be interpreted as saying that the density of the set $S := \{E \in \mathscr{E} : E(\mathbb{Q})_{\text{tors}} = 0\}$ is 1 (even though there do exist E with $E(\mathbb{Q})_{\text{tors}} \neq 0$).

2.2 Elliptic curves over local fields. Our model will be inspired by theorems and conjectures about the arithmetic of elliptic curves over $\mathbb{Q}$. But before studying elliptic curves over $\mathbb{Q}$, we should thoroughly understand elliptic curves over local fields.

Let $\mathbb{Q}_v$ be the completion of $\mathbb{Q}$ at a place v. There is a natural injective homomorphism $\operatorname{inv}\colon \mathrm{H}^2(\mathbb{Q}_v, \mathbb{G}_m) \to \mathbb{Q}/\mathbb{Z}$ that is an isomorphism if v is nonarchimedean.

Let E be an elliptic curve over $\mathbb{Q}_v$. Fix $n \geq 1$. Taking Galois cohomology in the exact sequence

$$0 \longrightarrow E[n] \longrightarrow E \xrightarrow{n} E \longrightarrow 0$$

yields a homomorphism $E(\mathbb{Q}_v)/nE(\mathbb{Q}_v) \to \mathrm{H}^1(\mathbb{Q}_v, E[n])$. Let W_v be its image. If v is a nonarchimedean place not dividing n and E has good reduction, then W_v equals the subgroup of unramified classes in $\mathrm{H}^1(\mathbb{Q}_v, E[n])$ [Poonen and Rains 2012, Proposition 4.13].

The theory of the Heisenberg group [Mumford 1991, pp. 44–46] yields an exact sequence

$$1 \longrightarrow \mathbb{G}_m \longrightarrow \mathfrak{H} \longrightarrow E[n] \longrightarrow 1,$$

which induces a map of sets

$$q_v\colon \mathrm{H}^1(\mathbb{Q}_v, E[n]) \longrightarrow \mathrm{H}^2(\mathbb{Q}_v, \mathbb{G}_m) \overset{\text{inv}}{\hookrightarrow} \mathbb{Q}/\mathbb{Z}.$$

It turns out that q_v is a quadratic form in the sense that $q_v(x+y) - q_v(x) - q_v(y)$ is bi-additive [Zarhin 1974b, §2]. Moreover, $q_v|_{W_v} = 0$ [O'Neil 2002, Proposition 2.3]. In fact, using Tate local duality one can show that W_v is a maximal isotropic subgroup of $\mathrm{H}^1(\mathbb{Q}_v, E[n])$ with respect to q_v [Poonen and Rains 2012, Proposition 4.11].

2.3 Selmer groups and Shafarevich–Tate groups. Now let E be an elliptic curve over $\mathbb{Q}$. Let $\mathbf{A} = \prod_v'(\mathbb{Q}_v, \mathbb{Z}_v)$ be the adèle ring of $\mathbb{Q}$; here v ranges over nontrivial places of $\mathbb{Q}$, Write $E(\mathbf{A})$ for $\prod_v E(\mathbb{Q}_v)/nE(\mathbb{Q}_v)$, and write $\mathrm{H}^1(\mathbf{A}, E[n])$ for the restricted product

$\prod'_v (\mathrm{H}^1(\mathbb{Q}_v, E[n]), W_v)$. We have a commutative diagram

$$\begin{array}{ccc} E(\mathbb{Q})/nE(\mathbb{Q}) & \longrightarrow & \mathrm{H}^1(\mathbb{Q}, E[n]) \\ \downarrow & & \downarrow \beta \\ E(\mathbf{A})/nE(\mathbf{A}) & \xrightarrow{\alpha} & \mathrm{H}^1(\mathbf{A}, E[n]). \end{array}$$

The n-Selmer group is defined by $\mathrm{Sel}_n E := \beta^{-1}(\operatorname{im} \alpha) \subseteq \mathrm{H}^1(\mathbb{Q}, E[n])$. (This is equivalent to the classical definition; we have only replaced $\prod_v \mathrm{H}^1(\mathbb{Q}_v, E[n])$ with a subgroup $\mathrm{H}^1(\mathbf{A}, E[n])$ into which α and β map.) The reason for defining $\mathrm{Sel}_n E$ is that it is a computable finite upper bound for (the image of) $E(\mathbb{Q})/nE(\mathbb{Q})$. On the other hand, the Shafarevich–Tate group is defined by

$$Ш = Ш(E) := \ker \left(\mathrm{H}^1(\mathbb{Q}, E) \to \prod_v \mathrm{H}^1(\mathbb{Q}_v, E) \right).$$

It is a torsion abelian group with an alternating pairing

$$[\ ,\ ] \colon Ш \times Ш \to \mathbb{Q}/\mathbb{Z}$$

defined by Cassels. Conjecturally, Ш is finite; in this case, $[\ ,\ ]$ is nondegenerate and #Ш is a square [Cassels 1962]. The definitions easily yield an exact sequence

$$0 \longrightarrow \frac{E(\mathbb{Q})}{nE(\mathbb{Q})} \longrightarrow \mathrm{Sel}_n E \longrightarrow Ш[n] \longrightarrow 0, \tag{1}$$

so $Ш[n]$ is measuring the difference between $\mathrm{Sel}_n E$ and the group $E(\mathbb{Q})/nE(\mathbb{Q})$ it is trying to approximate.

Each group in (1) decomposes according to the factorization of n into powers of distinct primes, so let us restrict to the case in which $n = p^e$ for some prime p and nonnegative integer e. Taking the direct limit over e yields an exact sequence

$$0 \longrightarrow E(\mathbb{Q}) \otimes \frac{\mathbb{Q}_p}{\mathbb{Z}_p} \longrightarrow \mathrm{Sel}_{p^\infty} E \longrightarrow Ш[p^\infty] \longrightarrow 0$$

of $\mathbb{Z}_p$-modules in which $\mathrm{Sel}_{p^\infty} E := \varinjlim \mathrm{Sel}_{p^e} E$ and $Ш[p^\infty] := \bigcup_{e \geq 0} Ш[p^e]$. Moreover, one can show that if $E(\mathbb{Q})[p] = 0$ (as holds for 100% of curves), then $\mathrm{Sel}_{p^e} E \to (\mathrm{Sel}_{p^\infty} E)[p^e]$ is an isomorphism (cf. Bhargava, Kane, Lenstra, Poonen, and Rains [2015, Proposition 5.9(b)]), so no information about the individual p^e-Selmer groups has been lost in passing to the limit.

2.4 The Selmer group as an intersection of maximal isotropic direct summands. If $\xi = (\xi_v) \in \mathrm{H}^1(\mathbf{A}, E[n])$, then for all but finitely many v we have $\xi_v \in W_v$ and hence $q_v(\xi_v) = 0$, so we may define $Q(\xi) := \sum_v q_v(\xi_v)$. This defines a quadratic form $Q\colon \mathrm{H}^1(\mathbf{A}, E[n]) \to \mathbb{Q}/\mathbb{Z}$.

Theorem 2.2.

(a) *Each of* $\operatorname{im}\alpha$ *and* $\operatorname{im}\beta$ *is a maximal isotropic subgroup of* $\mathrm{H}^1(\mathbf{A}, E[n])$ *with respect to* Q *[Poonen and Rains 2012, Theorem 4.14(a)].*

(b) *If* n *is prime or* $G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ *is surjective, then* β *is injective. (See Poonen and Rains [ibid., Proposition 3.3(e)] and Bhargava, Kane, Lenstra, Poonen, and Rains [2015, Proposition 6.1].)*

By definition, $\beta(\mathrm{Sel}_n\, E) = (\operatorname{im}\alpha) \cap (\operatorname{im}\beta)$. Thus, under either hypothesis in (b), $\mathrm{Sel}_n\, E$ is isomorphic to an intersection of maximal isotropic subgroups of $\mathrm{H}^1(\mathbf{A}, E[n])$.

Moreover, $\operatorname{im}\alpha$ is a direct summand of $\mathrm{H}^1(\mathbf{A}, E[n])$ [ibid., Corollary 6.8]. It is conjectured that $\operatorname{im}\beta$ is a direct summand as well, at least for asymptotically 100% of elliptic curves over $\mathbb{Q}$ [ibid., Conjecture 6.9], and it could hold for all of them.

2.5 The Birch and Swinnerton-Dyer conjecture. See Wiles [2006] for an introduction to the Birch and Swinnerton-Dyer conjecture more detailed than what we present here, and see Stein and Wuthrich [2013, Section 8] for some more recent advances towards it.

Let $E \in \mathscr{E}$. To E one can associate its L-function $L(E, s)$, a holomorphic function initially defined when $\operatorname{Re} s$ is sufficiently large, but known to extend to a holomorphic function on $\mathbb{C}$ (this is proved using the modularity of E). Just as the Dirichlet analytic class number formula expresses the residue at $s = 1$ of the Dedekind zeta function of a number field k in terms of the arithmetic of k, the Birch and Swinnerton-Dyer conjecture expresses the leading term in the Taylor expansion of $L(E, s)$ around $s = 1$ in terms of the arithmetic of E. We will state it only in the case that $\operatorname{rk} E(\mathbb{Q}) = 0$ since that is all that we will need. In addition to the quantities previously associated to E, we need

- the real period Ω, defined as the integral over $E(\mathbb{R})$ of a certain 1-form; and
- the Tamagawa number c_p for each finite prime p, a p-adic volume analogous to the real period.

Also define

$$Ш_0(E) := \begin{cases} \#Ш(E), & \text{if } \operatorname{rk} E(\mathbb{Q}) = 0; \\ 0, & \text{if } \operatorname{rk} E(\mathbb{Q}) > 0. \end{cases}$$

Conjecture 2.3 (The rank 0 part of the Birch and Swinnerton-Dyer conjecture). *If $E \in \mathscr{E}$, then*

$$L(E,1) = \frac{Ш_0 \, \Omega \, \prod_p c_p}{\#E(\mathbb{Q})_{\text{tors}}^2}. \tag{2}$$

Remark 2.4. In the case where the rank $r := \operatorname{rk} E(\mathbb{Q})$ is greater than 0, Conjecture 2.3 states only that $L(E,1) = 0$, whereas the full Birch and Swinnerton-Dyer conjecture predicts that $\operatorname{ord}_{s=1} L(E,s) = r$ and predicts the leading coefficient in the Taylor expansion of $L(E,s)$ at $s = 1$.

Let $H = \operatorname{ht} E$. Following Lang [1983] (see also Goldfeld and Szpiro [1995], de Weger [1998], Hindry [2007], Watkins [2008], and Hindry and Pacheco [2016]), we estimate the typical size of $Ш_0$ by estimating all the other quantities in (2) as $H \to \infty$; see Park, Poonen, Voight, and Wood [2016, Section 6] for details. The upshot is that if we average over E and ignore factors that are $H^{o(1)}$, then (2) simplifies to $1 \sim Ш_0 \, \Omega$ and we obtain $Ш_0 \sim \Omega^{-1} \sim H^{1/12}$. More precisely:

- $\prod_p c_p = H^{o(1)}$ [de Weger 1998, Theorem 3], [Hindry 2007, Lemma 3.5], [Watkins 2008, pp. 114–115], [Park, Poonen, Voight, and Wood 2016, Lemma 6.2.1];
- $\#E(\mathbb{Q})_{\text{tors}} \le 16$ [Mazur 1977];
- $\Omega = H^{-1/12+o(1)}$ [Hindry 2007, Lemma 3.7], [Park, Poonen, Voight, and Wood 2016, Corollary 6.1.3]; and
- the Riemann hypothesis for $L(E,s)$ implies that $L(E,1) \le H^{o(1)}$ [Iwaniec and Sarnak 2000, p. 713]. In fact, it is reasonable to expect $\operatorname*{Average}_{E \in \mathscr{E}_{\le H}} L(E,1) \asymp 1$. (The symbol $\asymp$ means that the left side is bounded above and below by positive constants times the right side.)

Thus we expect

$$\operatorname*{Average}_{E \in \mathscr{E}_{\le H}} Ш_0(E) = H^{1/12+o(1)} \tag{3}$$

as $H \to \infty$.

3 Modeling elliptic curves over $\mathbb{Q}$

3.1 Modeling the p-Selmer group. According to Theorem 2.2, $\operatorname{Sel}_p E$ is isomorphic to an intersection of maximal isotropic subspaces in an infinite-dimensional quadratic space over $\mathbb{F}_p$. So one might ask whether one could make sense of choosing maximal isotropic subspaces in an infinite-dimensional quadratic space at random, so that one could intersect two of them to obtain a space whose distribution is conjectured to be that of $\operatorname{Sel}_p E$. This can be done by equipping an infinite-dimensional quadratic space with a locally compact topology [Poonen and Rains 2012, Section 2], but the resulting distribution can be obtained more simply by working within a $2n$-dimensional quadratic space and taking the limit as $n \to \infty$. Now every nondegenerate $2n$-dimensional quadratic space with a maximal isotropic subspace is isomorphic to the quadratic space $V_n = (\mathbb{F}_p^{2n}, Q)$, where Q is the quadratic form

$$Q(x_1, \dots, x_n, y_1, \dots, y_n) := x_1 y_1 + \cdots + x_n y_n.$$

Therefore we conjecture that the distribution of $\dim_{\mathbb{F}_p} \operatorname{Sel}_p E$ as E varies over $\mathscr{E}$ equals the limit as $n \to \infty$ of the distribution of the dimension of the intersection of two maximal isotropic subspaces in V_n chosen uniformly at random from the finitely many possibilities. The limit exists and can be computed explicitly; this yields the formula on the right in the following:

Conjecture 3.1 (Poonen and Rains [ibid., Conjecture 1.1]). *For each $s \geq 0$, the density of $\{E \in \mathscr{E} : \dim_{\mathbb{F}_p} \operatorname{Sel}_p E = s\}$ equals*

$$\prod_{j \geq 0} (1 + p^{-j})^{-1} \prod_{j=1}^{s} \frac{p}{p^j - 1}. \tag{4}$$

Remark 3.2. Let E_d be the elliptic curve $dy^2 = x^3 - x$ over $\mathbb{Q}$. Heath-Brown [1993, 1994] proved that the density of integers d such that $\dim_{\mathbb{F}_2} \operatorname{Sel}_2 E_d - 2 = s$ equals

$$\prod_{j \geq 0} (1 + 2^{-j})^{-1} \prod_{j=1}^{s} \frac{2}{2^j - 1},$$

matching (4) for $p = 2$. (The -2 is there to remove the "causal" contribution to $\dim \operatorname{Sel}_2 E_d$ coming from $E_d(\mathbb{Q})[2]$.) As we have explained, this result is a natural consequence of the theory of Section 2.4, but in fact Heath-Brown's result came first and the theory was reverse engineered from it [Poonen and Rains 2012]! Heath-Brown's result was extended by Swinnerton-Dyer [2008] and Kane [2013] to the family of quadratic twists of any $E \in \mathscr{E}$ with $E[2] \subseteq E(\mathbb{Q})$ and no cyclic 4-isogeny.

3.2 Modeling the p^e-Selmer group.

If p is replaced by p^e, then we should replace $\mathbb{F}_p^{2n}$ by $V_n := ((\mathbb{Z}/p^e\mathbb{Z})^{2n}, Q)$. But now there are different types of maximal isotropic subgroups up to isomorphism. For example, if $e = 2$, then $(\mathbb{Z}/p^2\mathbb{Z})^n \times \{0\}^n$ and $(p\mathbb{Z}/p^2\mathbb{Z})^{2n}$

are both maximal isotropic subgroups; of these, only the first is a direct summand of V_n. In what follows, we will use only direct summands, for reasons to be explained at the end of this section.

Conjecture 3.3. *If we intersect two random maximal isotropic direct summands of $V_n := ((\mathbb{Z}/p^e\mathbb{Z})^{2n}, Q)$ and take the limit as $n \to \infty$ of the resulting distribution, we obtain the distribution of* $\operatorname{Sel}_{p^e} E$ *as E varies over $\mathscr{E}$.*

For $m \geq 1$, let $\sigma(m)$ denote the sum of the positive divisors of m. One can prove that the limit as $n \to \infty$ of the average size of the random intersection equals $\sigma(p^e)$, and there is an analogous result for positive integers m not of the form p^e [Bhargava, Kane, Lenstra, Poonen, and Rains 2015, Proposition 5.20]. This suggests the following:

Conjecture 3.4 (Poonen and Rains [2012, Conjecture 1(b)], Bhargava, Kane, Lenstra, Poonen, and Rains [2015, Section 5.7], Bhargava and Shankar [2013a, Conjecture 4])**.** *For each positive integer m,*

$$\operatorname*{Average}_{E \in \mathscr{E}} \# \operatorname{Sel}_m E = \sigma(m).$$

(The average is interpreted as the limit as $H \to \infty$ of the average over $\mathscr{E}_{\leq H}$.)

One could similarly compute the higher moments of the conjectural distribution; see Poonen and Rains [2012, Proposition 2.22(a)] and Bhargava, Kane, Lenstra, Poonen, and Rains [2015, Section 5.5].

There are several reasons why insisting upon direct summands in Conjecture 3.3 seems right:

- Conjecturally, both of the maximal isotropic subgroups arising in the arithmetic of the elliptic curve *are* direct summands: see the last paragraph of Section 2.4.
- Requiring direct summands is essentially the only way to make the model for $\operatorname{Sel}_{p^e} E$ consistent with the model for $\operatorname{Sel}_p E$, given that $\operatorname{Sel}_p E \simeq (\operatorname{Sel}_{p^e} E)[p]$ for 100% of curves [ibid., Remark 6.12].
- It leads to Conjecture 3.4, which has been proved for $m \leq 5$ [Bhargava and Shankar 2015a,b, 2013a,b].

3.3 Modeling the p^∞-Selmer group and the Shafarevich–Tate group. Choosing a maximal isotropic direct summand of $((\mathbb{Z}/p^e\mathbb{Z})^{2n}, Q)$ compatibly for all e is equivalent to choosing a maximal isotropic direct summand of the quadratic $\mathbb{Z}_p$-module $V_n := (\mathbb{Z}_p^{2n}, Q)$. This observation will lead us to a process that models $\operatorname{Sel}_{p^e} E$ for all e simultaneously, or equivalently, that models $\operatorname{Sel}_{p^\infty} E$ directly. To simplify notation, for any

$\mathbb{Z}_p$-module X, let X' denote $X \otimes \frac{\mathbb{Q}_p}{\mathbb{Z}_p}$; if X is a $\mathbb{Z}_p$-submodule of V_n, then X' is a $\mathbb{Z}_p$-submodule of V_n'.

Now choose maximal isotropic direct summands Z and W of V_n with respect to the measure arising from taking the inverse limit over e of the uniform measure on the set of maximal isotropic direct summands of $(\mathbb{Z}/p^e\mathbb{Z})^{2n}$ [Bhargava, Kane, Lenstra, Poonen, and Rains 2015, Sections 2 and 4]; then we conjecture that the limiting distribution of $Z' \cap W'$ as $n \to \infty$ equals the distribution of $\operatorname{Sel}_{p^\infty} E$ as E varies over $\mathscr{E}$. Again, the point is that this limiting distribution is compatible with the previously conjectured distribution for $\operatorname{Sel}_{p^e} E$ for each nonnegative integer e, and the conjecture for $\operatorname{Sel}_{p^e} E$ was based on *theorems* about Selmer groups of elliptic curves (see Section 2.4).

Even better, using the same ingredients, we can model $\operatorname{rk} E(\mathbb{Q})$ and $Ш[p^\infty]$ at the same time:

Conjecture 3.5 (Bhargava, Kane, Lenstra, Poonen, and Rains [ibid., Conjecture 1.3])**.** *If we choose maximal isotropic direct summands Z and W of $(\mathbb{Z}_p^{2n}, Q)$ at random as above, and we define*

$$R := (Z \cap W)', \qquad S := Z' \cap W', \qquad T := S/R,$$

then the limit as $n \to \infty$ of the distribution of the exact sequence

$$0 \longrightarrow R \longrightarrow S \longrightarrow T \longrightarrow 0$$

equals the distribution of the sequence

$$0 \longrightarrow E(\mathbb{Q}) \otimes \frac{\mathbb{Q}_p}{\mathbb{Z}_p} \longrightarrow \operatorname{Sel}_{p^\infty} E \longrightarrow Ш[p^\infty] \longrightarrow 0$$

as E varies over $\mathscr{E}$.

There are several pieces of indirect evidence for the rank and Ш predictions of Conjecture 3.5:

- Each of R and $E(\mathbb{Q}) \otimes \frac{\mathbb{Q}_p}{\mathbb{Z}_p}$ is isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^r$ for some nonnegative integer r, called the $\mathbb{Z}_p$-corank of the module.
- The $\mathbb{Z}_p$-corank of R is 0 or 1, with probability $1/2$ each [ibid., Proposition 5.6]. Likewise, a variant of a conjecture of Goldfeld (see Goldfeld [1979, Conjecture B] and Katz and Sarnak [1999a,b]) predicts that $\operatorname{rk} E(\mathbb{Q})$ (which equals the $\mathbb{Z}_p$-corank of $E(\mathbb{Q}) \otimes \frac{\mathbb{Q}_p}{\mathbb{Z}_p}$) is 0, 1, ≥ 2 with densities $1/2$, $1/2$, 0, respectively.
- The group T is finite and carries a nondegenerate alternating pairing with values in $\mathbb{Q}_p/\mathbb{Z}_p$, just as $Ш[p^\infty]$ conjecturally does (the p-part of the Cassels pairing). In particular, $\#T$ is a square.

- [Smith 2017] has proved a result analogous to Conjecture 3.5 for the family of quadratic twists of any $E \in \mathscr{E}$ with $E[2] \subseteq E(\mathbb{Q})$ and no cyclic 4-isogeny.

Further evidence is that there are in fact *three* distributions that have been conjectured to be the distribution of $Ш[p^\infty]$ as E varies over rank r elliptic curves, and these three distributions coincide [Bhargava, Kane, Lenstra, Poonen, and Rains 2015, Theorems 1.6(c) and 1.10(b)]. This is so even in the cases with $r \geq 2$, which conjecturally occur with density 0. For a fixed nonnegative integer r, the three distributions are as follows:

1. A distribution defined by Delaunay [2001, 2007] and Delaunay and Jouhet [2014], who adapted the Cohen–Lenstra heuristics for class groups [Cohen and Lenstra 1984].

2. The limit as $n \to \infty$ of the distribution of $T := (Z' \cap W')/(Z \cap W)'$ when (Z, W) is sampled from the set of pairs of maximal isotropic direct summands of $(\mathbb{Z}_p^{2n}, Q)$ satisfying $\operatorname{rk}_{\mathbb{Z}_p}(Z \cap W) = r$. (This set of pairs is the set of $\mathbb{Z}_p$-points of a scheme of finite type, so it carries a natural measure [Bhargava, Kane, Lenstra, Poonen, and Rains 2015, Sections 2 and 4].)

3. The limit as $n \to \infty$ through integers of the same parity as r of the distribution of $(\operatorname{coker} A)_{\text{tors}}$ when A is sampled from the space of matrices in $\mathrm{M}_n(\mathbb{Z}_p)$ satisfying $A^T = -A$ and $\operatorname{rk}_{\mathbb{Z}_p}(\ker A) = r$; here $\ker A$ and $\operatorname{coker} A$ are defined by viewing A as a $\mathbb{Z}_p$-linear homomorphism $\mathbb{Z}_p^n \to \mathbb{Z}_p^n$.

The last of these is inspired by the theorem of Friedman and Washington [1989] that for each odd prime p, the limit as $n \to \infty$ of the distribution coker A for $A \in \mathrm{M}_n(\mathbb{Z}_p)$ chosen at random with respect to Haar measure equals the distribution conjectured by Cohen and Lenstra to be the distribution of the p-primary part of the class group of a varying imaginary quadratic field.

3.4 Modeling the rank of an elliptic curve. In the previous section, we saw in the third construction that conditioning on $\operatorname{rk}_{\mathbb{Z}_p}(\ker A) = r$ yields the conjectural distribution of $Ш[p^\infty]$ for rank r curves. The simplest possible explanation for this would be that sampling A at random from $\mathrm{M}_n(\mathbb{Z}_p)_{\text{alt}} := \{A \in \mathrm{M}_n(\mathbb{Z}_p) : A^T = -A\}$ *without* conditioning on $\operatorname{rk}_{\mathbb{Z}_p}(\ker A)$ caused $\operatorname{rk}_{\mathbb{Z}_p}(\ker A)$ to be distributed like the rank of an elliptic curve.

What is the distribution of $\operatorname{rk}_{\mathbb{Z}_p}(\ker A)$? If n is even, then the locus in $\mathrm{M}_n(\mathbb{Z}_p)_{\text{alt}}$ defined by $\det A = 0$ is the set of $\mathbb{Z}_p$-points of a hypersurface, which has Haar measure 0, so $\operatorname{rk}_{\mathbb{Z}_p}(\ker A) = 0$ with probability 1. If n is odd, however, then $\operatorname{rk}_{\mathbb{Z}_p}(\ker A)$ cannot be 0, because $n - \operatorname{rk}_{\mathbb{Z}_p}(\ker A)$ is the rank of A, which is even for an alternating matrix. For n odd, it turns out that $\operatorname{rk}_{\mathbb{Z}_p}(\ker A) = 1$ with probability 1. If we imagine that n was chosen large and with random parity, then the result is that $\operatorname{rk}_{\mathbb{Z}_p}(\ker A)$ is 0 or 1, with probability 1/2 each. This result agrees with the variant of Goldfeld's conjecture mentioned above.

This model cannot, however, distinguish the relative frequencies of curves of various ranks ≥ 2, because in the model the event $\operatorname{rk}_{\mathbb{Z}_p}(\ker A) \geq 2$ occurs with probability 0.

Therefore we propose a refined model in which instead of sampling A from $\mathrm{M}_n(\mathbb{Z}_p)_{\mathrm{alt}}$, we sample A from the set $\mathrm{M}_n(\mathbb{Z})_{\mathrm{alt},\leq X}$ of matrices in $\mathrm{M}_n(\mathbb{Z})_{\mathrm{alt}}$ with entries bounded by a number X *depending on the height H of the elliptic curve being modeled*, tending to ∞ as $H \to \infty$. This way, for elliptic curves of a given height H, the model predicts a potentially positive but diminishing probability of each rank ≥ 2 (the probability that an integer point in a box lies on a certain subvariety), and we can quantify the rate at which this probability tends to 0 as $H \to \infty$ in order to count curves of height up to H having each given rank. In fact, we let n grow with H as well.

Here, more precisely, is the refined model. To model an elliptic curve E of height H, using functions $\eta(H)$ and $X(H)$ to be specified later, we do the following:

1. Choose n to be an integer of size about $\eta(H)$ of random parity (e.g., we could choose n uniformly at random from $\{\lceil \eta(H) \rceil, \lceil \eta(H) \rceil + 1\}$).

2. Choose $A_E \in \mathrm{M}_n(\mathbb{Z})_{\mathrm{alt},\leq X(H)}$ uniformly at random, independently for each E.

3. Define random variables $\text{Ш}'_E := (\operatorname{coker} A)_{\mathrm{tors}}$ and $\operatorname{rk}'_E := \operatorname{rk}_{\mathbb{Z}}(\ker A)$.

Think of $\text{Ш}'_E$ as the "pseudo-Shafarevich–Tate group" of E and rk'_E as the "pseudo-rank" of E; their behavior is intended to model the actual Ш and rank.

To complete the description of the model, we must specify the functions $\eta(H)$ and $X(H)$. We do this by asking "How large is Ш_0 on average?", both in the model and in reality. Recall from (3) that we expect

(5) $$\operatorname*{Average}_{E \in \mathscr{E}_{\leq H}} \text{Ш}_0(E) = H^{1/12+o(1)}.$$

Define

$$\text{Ш}'_{E,0} := \begin{cases} \#\text{Ш}'_E, & \text{if } \operatorname{rk}'_E = 0; \\ 0, & \text{if } \operatorname{rk}'_E > 0. \end{cases}$$

Using that the determinant of an $n \times n$ matrix is given by a polynomial of degree n in the entries, we can prove that

(6) $$\operatorname*{Average}_{E \in \mathscr{E}_{\leq H}} \text{Ш}'_{E,0} = X(H)^{\eta(H)(1+o(1))},$$

assuming that $\eta(H)$ does not grow too quickly with H. Comparing (5) and (6) suggests choosing $\eta(H)$ and $X(H)$ so that $X(H)^{\eta(H)} = H^{1/12+o(1)}$. We assume this from now on. It turns out that we will not need to know any more about $\eta(H)$ and $X(H)$ than this.

3.5 Consequences of the model. To see what distribution of ranks is predicted by the refined model, we must calculate the distribution of ranks of alternating matrices whose entries are integers with bounded absolute value; the relevant result, whose proof is adapted from Eskin and Katznelson [1995], is the following:

Theorem 3.6 (cf. Park, Poonen, Voight, and Wood [2016, Theorem 9.1.1]). *If $1 \le r \le n$ and $n - r$ is even, and $A \in \mathrm{M}_n(\mathbb{Z})_{\mathrm{alt}, \le X}$ is chosen uniformly at random, then*

$$\mathrm{Prob}(\mathrm{rk}(\ker A) \ge r) \asymp_n (X^n)^{-(r-1)/2}.$$

(The subscript n on the symbol $\asymp$ means that the implied constants depend on n.)

Theorem 3.6 implies that for fixed $r \ge 1$ and $E \in \mathscr{E}$ of height H,

$$\mathrm{Prob}(\mathrm{rk}'_E \ge r) = (X(H)^{\eta(H)})^{-(r-1)/2+o(1)} = H^{-(r-1)/24+o(1)}. \tag{7}$$

Using this, and the fact $\#\mathscr{E}_{\le H} \asymp H^{5/6} = H^{20/24}$ (Proposition 2.1), we can now sum (7) over $E \in \mathscr{E}_{\le H}$ to prove the following theorem about our model:

Theorem 3.7 (Park, Poonen, Voight, and Wood [ibid., Theorem 7.3.3]). *The following hold with probability* 1*:*

$$\begin{aligned}
\#\{E \in \mathscr{E}_{\le H} : \mathrm{rk}'_E = 0\} &= H^{20/24+o(1)} \\
\#\{E \in \mathscr{E}_{\le H} : \mathrm{rk}'_E = 1\} &= H^{20/24+o(1)} \\
\#\{E \in \mathscr{E}_{\le H} : \mathrm{rk}'_E \ge 2\} &= H^{19/24+o(1)} \\
\#\{E \in \mathscr{E}_{\le H} : \mathrm{rk}'_E \ge 3\} &= H^{18/24+o(1)} \\
&\vdots \\
\#\{E \in \mathscr{E}_{\le H} : \mathrm{rk}'_E \ge 20\} &= H^{1/24+o(1)} \\
\#\{E \in \mathscr{E}_{\le H} : \mathrm{rk}'_E \ge 21\} &\le H^{o(1)}, \\
\#\{E \in \mathscr{E} : \mathrm{rk}'_E > 21\} &\text{ is finite.}
\end{aligned}$$

This suggests the conjecture that the same statements hold for the *actual* ranks of elliptic curves over $\mathbb{Q}$. In particular, we conjecture that $\mathrm{rk}\, E(\mathbb{Q})$ is uniformly bounded, bounded by the maximum of the ranks of the conjecturally finitely many elliptic curves of rank > 21.

Remark 3.8. Elkies [2006] has found infinitely many elliptic curves over $\mathbb{Q}$ of rank ≥ 19, and one of rank ≥ 28; these have remained the records since 2006.

4 Generalizations

4.1 Elliptic curves over global fields. What about elliptic curves over other global fields K? Let $\mathscr{E}_K$ be a set of representatives for the isomorphism classes of elliptic curves over K. Let $B_K := \limsup_{E \in \mathscr{E}_K} \operatorname{rk} E(K)$. For example, the conjecture suggested by Theorem 3.7 predicts that $20 \le B_{\mathbb{Q}} \le 21$.

Theorem 4.1 (Tèĭt and Šafarevič [1967], Ulmer [2002]). *If K is a global function field, then $B_K = \infty$.*

Even for number fields, B_K can be arbitrarily large (but maybe still always finite):

Theorem 4.2 (Park, Poonen, Voight, and Wood [2016, Theorem 12.4.2]). *There exist number fields K of arbitrarily high degree such that $B_K \ge [K : \mathbb{Q}]$.*

Examples of number fields K for which B_K is large include fields in anticyclotomic towers and certain multiquadratic fields; see Park, Poonen, Voight, and Wood [ibid., Section 12.4].

A naive adaptation of our heuristic (see Park, Poonen, Voight, and Wood [ibid., Sections 12.2 and 12.3]) would suggest $20 \le B_K \le 21$ for every global field K, but Theorems 4.1 and 4.2 contradict this conclusion. Our rationalization of this is that the elliptic curves of high rank in Theorems 4.1 and 4.2 are special in that they are definable over a proper subfield of K, and these special curves exhibit arithmetic phenomena that our model does not take into account. To exclude these curves, let $\mathscr{E}_K^\circ$ be the set of $E \in \mathscr{E}_K$ such that E is not a base change of a curve from a proper subfield, and let $B_K^\circ := \limsup_{E \in \mathscr{E}_K^\circ} \operatorname{rk} E(K)$. It is possible that $B_K^\circ < \infty$ for every global field K.

Remark 4.3. On the other hand, it is not true that $B_K^\circ \le 21$ for all number fields, as we now explain. Shioda [1992] proved that $y^2 = x^3 + t^{360} + 1$ has rank 68 over $\mathbb{C}(t)$. In fact, it has rank 68 also over $K(t)$ for a suitable number field K. For this K, specialization yields infinitely many elliptic curves in $\mathscr{E}_K^\circ$ of rank ≥ 68. Thus $B_K^\circ \ge 68$. See Park, Poonen, Voight, and Wood [2016, Remark 12.3.1] for details.

4.2 Abelian varieties.

Question 4.4. For abelian varieties A over number fields K, is there a bound on $\operatorname{rk} A(K)$ depending only on $\dim A$ and $[K : \mathbb{Q}]$?

By restriction of scalars, we can reduce to the case $K = \mathbb{Q}$ at the expense of increasing the dimension. By "Zarhin's trick" that $A^4 \times (A^\vee)^4$ is principally polarized [Zarhin 1974a], we can reduce to the case that A is principally polarized, again at the expense of increasing the dimension. For fixed $g \geq 0$, one can write down a family of projective varieties including all g-dimensional principally polarized abelian varieties over $\mathbb{Q}$, probably with each isomorphism class represented infinitely many times. We can assume that each abelian variety A is defined by a system of homogeneous polynomials with integer coefficients, in which the number of variables, the number of polynomials, and their degrees are bounded in terms of g. Define the height of A to be the maximum of the absolute values of the coefficients. Then the number of g-dimensional principally polarized abelian varieties over $\mathbb{Q}$ of height $\leq H$ is bounded by a polynomial in H. If there is a model involving a pseudo-rank rk'_A whose probability of exceeding r gets divided by at least a fixed fractional power of H each time r is incremented by 1, as we had for elliptic curves, then the pseudo-ranks are bounded with probability 1. This might suggest a positive answer to Question 4.4, though the evidence is much flimsier than in the case of elliptic curves.

Acknowledgments. I thank Nicolas Billerey, Serge Cantat, Andrew Granville, Eric Rains, Michael Stoll, and John Voight for comments.

References

Jennifer S. Balakrishnan, Wei Ho, Nathan Kaplan, Simon Spicer, William Stein, and James Weigandt (2016). "Databases of elliptic curves ordered by height and distributions of Selmer groups and ranks". *LMS J. Comput. Math.* 19.suppl. A, pp. 351–370. MR: 3540965 (cit. on p. 418).

Manjul Bhargava, Daniel M. Kane, Hendrik W. Lenstra Jr., Bjorn Poonen, and Eric Rains (2015). "Modeling the distribution of ranks, Selmer groups, and Shafarevich–Tate groups of elliptic curves". *Camb. J. Math.* 3.3, pp. 275–321. MR: 3393023 (cit. on pp. 417, 420, 421, 424–426).

Manjul Bhargava and Arul Shankar (Dec. 27, 2013a). *The average number of elements in the 4-Selmer groups of elliptic curves is 7*. Preprint. arXiv: 1312.7333 (cit. on p. 424).

– (Dec. 30, 2013b). *The average size of the 5-Selmer group of elliptic curves is 6, and the average rank is less than 1*. Preprint. arXiv: 1312.7859 (cit. on p. 424).

– (2015a). "Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves". *Ann. of Math. (2)* 181.1, pp. 191–242. MR: 3272925 (cit. on p. 424).

– (2015b). "Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0". *Ann. of Math. (2)* 181.2, pp. 587–621. MR: 3275847 (cit. on p. 424).

Armand Brumer (1992). "The average rank of elliptic curves. I". *Invent. Math.* 109.3, pp. 445–472. MR: 1176198 (cit. on pp. 417, 418).

J. W. S. Cassels (1962). "Arithmetic on curves of genus 1. IV. Proof of the Hauptvermutung". *J. Reine Angew. Math.* 211, pp. 95–112. MR: 0163915 (cit. on p. 420).

– (1966). "Diophantine equations with special reference to elliptic curves". *J. London Math. Soc.* 41, pp. 193–291. MR: 0199150 (cit. on p. 417).

H. Cohen and Hendrik W. Lenstra Jr. (1984). "Heuristics on class groups of number fields". In: *Number theory, Noordwijkerhout 1983*. Vol. 1068. Lecture Notes in Math. Berlin: Springer, pp. 33–62. MR: 0756082 (cit. on p. 426).

Christophe Delaunay (2001). "Heuristics on Tate–Shafarevitch groups of elliptic curves defined over $\mathbb{Q}$". *Experiment. Math.* 10.2, pp. 191–196. MR: 1837670 (cit. on p. 426).

– (2007). "Heuristics on class groups and on Tate–Shafarevich groups: the magic of the Cohen–Lenstra heuristics". In: *Ranks of elliptic curves and random matrix theory*. Vol. 341. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press, pp. 323–340. MR: 2322355 (cit. on p. 426).

Christophe Delaunay and Frédéric Jouhet (2014). "p^ℓ-torsion points in finite abelian groups and combinatorial identities". *Adv. Math.* 258, pp. 13–45. MR: 3190422 (cit. on p. 426).

Noam D. Elkies (May 3, 2006). *$\mathbb{Z}^{28}$ in $E(\mathbb{Q})$, etc.* Email to the number theory mailing list at NMBRTHRY@LISTSERV.NODAK.EDU (cit. on p. 428).

Alex Eskin and Yonatan R. Katznelson (1995). "Singular symmetric matrices". *Duke Math. J.* 79.2, pp. 515–547. MR: 1344769 (cit. on p. 428).

David W. Farmer, S. M. Gonek, and C. P. Hughes (2007). "The maximum size of L-functions". *J. Reine Angew. Math.* 609, pp. 215–236. MR: 2350784 (cit. on p. 417).

Eduardo Friedman and Lawrence C. Washington (1989). "On the distribution of divisor class groups of curves over a finite field". In: *Théorie des nombres, Quebec 1987*. Berlin: de Gruyter, pp. 227–239. MR: 1024565 (cit. on p. 426).

Dorian Goldfeld (1979). "Conjectures on elliptic curves over quadratic fields". In: *Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979)*. Vol. 751. Lecture Notes in Math. Berlin: Springer, pp. 108–118. MR: 0564926 (cit. on p. 425).

Dorian Goldfeld and Lucien Szpiro (1995). "Bounds for the order of the Tate-Shafarevich group". *Compositio Math.* 97.1-2. Special issue in honour of Frans Oort, pp. 71–87. MR: 1355118 (cit. on p. 422).

D. R. Heath-Brown (1993). "The size of Selmer groups for the congruent number problem". *Invent. Math.* 111.1, pp. 171–195. MR: 1193603 (cit. on p. 423).

– (1994). "The size of Selmer groups for the congruent number problem. II". *Invent. Math.* 118.2. With an appendix by P. Monsky, pp. 331–370. MR: 1292115 (cit. on p. 423).

Marc Hindry (2007). "Why is it difficult to compute the Mordell–Weil group?" In: *Diophantine geometry*. Vol. 4. CRM Series. Ed. Norm., Pisa, pp. 197–219. MR: 2349656 (cit. on p. 422).

Marc Hindry and Amílcar Pacheco (2016). "An analogue of the Brauer–Siegel theorem for abelian varieties in positive characteristic". *Mosc. Math. J.* 16.1, pp. 45–93. MR: 3470576 (cit. on p. 422).

Taira Honda (1960). "Isogenies, rational points and section points of group varieties". *Japan. J. Math.* 30, pp. 84–101. MR: 0155828 (cit. on p. 417).

H. Iwaniec and P. Sarnak (2000). "Perspectives on the analytic theory of L-functions". *Geom. Funct. Anal.* Special Volume. GAFA 2000 (Tel Aviv, 1999), pp. 705–741. MR: 1826269 (cit. on p. 422).

Daniel M. Kane (2013). "On the ranks of the 2-Selmer groups of twists of a given elliptic curve". *Algebra Number Theory* 7.5, pp. 1253–1279. MR: 3101079 (cit. on p. 423).

Nicholas M. Katz and Peter Sarnak (1999a). *Random matrices, Frobenius eigenvalues, and monodromy*. Vol. 45. American Mathematical Society Colloquium Publications. Amer. Math. Soc., pp. xii+419. MR: 1659828 (cit. on p. 425).

– (1999b). "Zeroes of zeta functions and symmetry". *Bull. Amer. Math. Soc. (N.S.)* 36.1, pp. 1–26. MR: 1640151 (cit. on p. 425).

William E. Lang (1983). "On Enriques surfaces in characteristic p. I". *Math. Ann.* 265.1, pp. 45–65. MR: 0719350 (cit. on p. 422).

B. Mazur (1977). "Modular curves and the Eisenstein ideal". *Inst. Hautes Études Sci. Publ. Math.* 47, 33–186 (1978). MR: 0488287 (cit. on p. 422).

Jean-François Mestre (1982). "Construction d'une courbe elliptique de rang $\geq$ 12". French, with English summary. *C. R. Acad. Sci. Paris Sér. I Math.* 295.12, pp. 643–644. MR: 0688896 (cit. on p. 417).

– (1986). "Formules explicites et minorations de conducteurs de variétés algébriques". French. *Compositio Math.* 58.2, pp. 209–232. MR: 0844410 (cit. on p. 417).

L. J. Mordell (1922). "On the rational solutions of the indeterminate equations of the third and fourth degrees". *Proc. Cambridge Phil. Soc.* 21, pp. 179–192 (cit. on p. 417).

David Mumford (1991). *Tata lectures on theta. III*. Vol. 97. Progress in Mathematics. With the collaboration of Madhav Nori and Peter Norman. Birkhäuser Boston Inc., pp. viii+202. MR: 1116553 (cit. on p. 419).

Catherine O'Neil (2002). "The period-index obstruction for elliptic curves". *J. Number Theory* 95.2. Erratum in *J. Number Theory* 109 (2004), no. 2, 390, pp. 329–339. MR: 1924106 (cit. on p. 419).

Jennifer Park, Bjorn Poonen, John Voight, and Melanie Matchett Wood (Feb. 3, 2016). *A heuristic for boundedness of ranks of elliptic curves*. Preprint, to appear in *J. Europ. Math. Soc.* arXiv: 1602.01431 (cit. on pp. 417, 422, 428, 429).

H. Poincaré (1901). "Sur les propriétés arithmétiques des courbes algébriques". *J. Pures Appl. Math. (5)* 7, pp. 161–234 (cit. on p. 417).

Henri Poincaré (1950). *Œuvres d'Henri Poincaré*. Vol. 5. Edited by Albert Châtelet. Paris: Gauthier-Villars. MR: 0044457 (cit. on p. 417).

Bjorn Poonen and Eric Rains (2012). "Random maximal isotropic subspaces and Selmer groups". *J. Amer. Math. Soc.* 25.1, pp. 245–269. MR: 2833483 (cit. on pp. 417, 419, 421, 423, 424).

Karl Rubin and Alice Silverberg (2000). "Ranks of elliptic curves in families of quadratic twists". *Experiment. Math.* 9.4, pp. 583–590. MR: 1806293 (cit. on p. 418).

Tetsuji Shioda (1992). "Some remarks on elliptic curves over function fields". *Astérisque* 209. Journées Arithmétiques, 1991 (Geneva), pp. 12, 99–114. MR: 1211006 (cit. on p. 429).

Joseph H. Silverman (2009). *The arithmetic of elliptic curves*. 2nd ed. Vol. 106. Graduate Texts in Mathematics. Springer, Dordrecht, pp. xx+513. MR: 2514094 (cit. on p. 417).

Alexander Smith (June 7, 2017). *2^∞-Selmer groups, 2^∞-class groups, and Goldfeld's conjecture*. Preprint. arXiv: 1702.02325 (cit. on p. 426).

William Stein and Christian Wuthrich (2013). "Algorithms for the arithmetic of elliptic curves using Iwasawa theory". *Math. Comp.* 82.283, pp. 1757–1792. MR: 3042584 (cit. on p. 421).

Peter Swinnerton-Dyer (2008). "The effect of twisting on the 2-Selmer group". *Math. Proc. Cambridge Philos. Soc.* 145.3, pp. 513–526. MR: 2464773 (cit. on p. 423).

John T. Tate (1974). "The arithmetic of elliptic curves". *Invent. Math.* 23, pp. 179–206. MR: 0419359 (cit. on p. 417).

D. T. Tèĭt and I. R. Šafarevič (1967). "The rank of elliptic curves". Russian. *Dokl. Akad. Nauk SSSR* 175, pp. 770–773. MR: 0237508 (cit. on p. 429).

Douglas Ulmer (2002). "Elliptic curves with large rank over function fields". *Ann. of Math. (2)* 155.1, pp. 295–315. MR: 1888802 (cit. on pp. 417, 429).

Mark Watkins (2008). "Some heuristics about elliptic curves". *Experiment. Math.* 17.1, pp. 105–125. arXiv: math/0608766. MR: 2410120 (cit. on p. 422).

– (Aug. 20, 2015). *A discursus on 21 as a bound for ranks of elliptic curves over $\mathbf{Q}$, and sundry related topics*. http://magma.maths.usyd.edu.au/~watkins/papers/DISCURSUS.pdf (cit. on p. 418).

Mark Watkins, Stephen Donnelly, Noam D. Elkies, Tom Fisher, Andrew Granville, and Nicholas F. Rogers (2014). "Ranks of quadratic twists of elliptic curves". English, with English and French summaries. *Publ. math. de Besançon* 2014/2, pp. 63–98. MR: 3381037 (cit. on p. 418).

Benjamin M. M. de Weger (1998). "$A + B = C$ and big Ш's". English. *Quart. J. Math. Oxford Ser. (2)* 49.193, pp. 105–128. MR: 1617347 (cit. on p. 422).

Andrew Wiles (2006). "The Birch and Swinnerton-Dyer conjecture". In: *The millennium prize problems*. Clay Math. Inst., pp. 31–41. MR: 2238272 (cit. on p. 421).

Ju. G. Zarhin (1974a). "A remark on endomorphisms of abelian varieties over function fields of finite characteristic". Russian. *Izv. Akad. Nauk SSSR Ser. Mat.* 38, pp. 471–474. MR: 0354689 (cit. on p. 429).

Ju. G. Zarhin (1974b). “Noncommutative cohomology and Mumford groups”. Russian. *Mat. Zametki* 15, pp. 415–419. MR: 0354612 (cit. on p. 419).

Received 2017-11-30.

Bjorn Poonen
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139-4307
USA
poonen@math.mit.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (433–452)

POTENTIAL AUTOMORPHY OF $\widehat{G}$-LOCAL SYSTEMS

Jack Thorne

Abstract

Vincent Lafforgue has recently made a spectacular breakthrough in the setting of the global Langlands correspondence for global fields of positive characteristic, by constructing the 'automorphic–to–Galois' direction of the correspondence for an arbitrary reductive group G. We discuss a result that starts with Lafforgue's work and proceeds in the opposite ('Galois–to–automorphic') direction.

1 Introduction

Let $\mathbb{F}_q$ be a finite field, and let G be a split reductive group over $\mathbb{F}_q$. Let X be a smooth projective connected curve over $\mathbb{F}_q$, and let K be its function field. Let $\mathbb{A}_K$ denote the ring of adèles of K. Automorphic forms on G are locally constant functions $f : G(\mathbb{A}_K) \to \overline{\mathbb{Q}}$ which are invariant under left translation by the discrete group $G(K) \subset G(\mathbb{A}_K)$.[1] The space of automorphic forms is a representation of $G(\mathbb{A}_K)$, and its irreducible constituents are called automorphic representations.

Let $\ell \nmid q$ be a prime. According to the Langlands conjectures, any automorphic representation π of $G(\mathbb{A}_K)$ should give rise to a continuous representation $\rho(\pi) : \pi_1^{\text{ét}}(U) \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ (for some open subscheme $U \subset X$). This should be compatible with the (known) unramified local Langlands correspondence, which describes the pullback of $\rho(\pi)$ to $\pi_1^{\text{ét}}(\mathbb{F}_{q_v})$ for every closed point $v : \operatorname{Spec} \mathbb{F}_{q_v} \hookrightarrow U$ in terms of the components π_v of a factorization $\pi = \otimes'_v \pi_v$ into representations of the local groups $G(K_v)$.

Vincent Lafforgue has given an amazing construction of the representation $\rho(\pi)$, which crystallizes and removes many of the ambiguities in this picture in a beautiful way. We give a brief description of this work below.

Our main goal in this article is to describe a work due to Gebhard Böckle, Michael Harris, Chandrashekhar Khare, and myself, where we establish a partial converse to this

MSC2010: 11F80.

[1] This is not the full definition. In the rest of this article we consider only cuspidal automorphic forms, which are defined precisely below.

result Böckle, Harris, Khare, and J. A. Thorne [n.d.]. We restrict to representations $\rho : \pi_1^{\text{ét}}(X) \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ of Zariski dense image, and show that any such representation is *potentially automorphic*, in the sense that there exists a Galois cover $Y \to X$ such that the pullback of ρ to $\pi_1^{\text{ét}}(Y)$ is contained in the image of Lafforgue's construction. We will guide the reader through the context surrounding this result, and discuss some interesting open questions that are suggested by our methods.

2 Review of the case $G = \mathrm{GL}_n$

We begin by describing what is known about the Langlands conjectures in the setting of the general linear group $G = \mathrm{GL}_n$. In this case very complete results were obtained by L. Lafforgue [2002]. As in the introduction, we write $\mathbb{F}_q$ for the finite field with q elements, and let X be a smooth, projective, connected curve over $\mathbb{F}_q$. The Langlands correspondence predicts a relation between representations of the absolute Galois group of K and automorphic representations of the group $\mathrm{GL}_n(\mathbb{A}_K)$. We now describe each of these in turn.

Let K^s be a fixed separable closure of K. We write $\Gamma_K = \mathrm{Gal}(K^s/K)$ for the absolute Galois group of K, relative to K^s. It is a profinite group. If $S \subset X$ is a finite set of closed points, then we write $K_S \subset K^s$ for the maximal extension of K which is unramified outside S, and $\Gamma_{K,S} = \mathrm{Gal}(K_S/K)$ for its Galois group. This group has a geometric interpretation: if we set $U = X - S$, and write $\overline{\eta}$ for the geometric generic point of U corresponding to K^s, then there is a canonical identification $\Gamma_{K,S} \cong \pi_1^{\text{ét}}(U, \overline{\eta})$ of the Galois group with the étale fundamental group of the open curve U.

Fix a prime $\ell \nmid q$ and a continuous character $\omega : \Gamma_K \to \overline{\mathbb{Q}}_\ell^\times$ of finite order. If $n \geq 1$ is an integer, then we write $\mathrm{Gal}_{n,\omega}$ for the set of conjugacy classes of continuous representations $\rho : \Gamma_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$ with the following properties:

1. ρ factors through a quotient $\Gamma_K \to \Gamma_{K,S}$, for some finite subscheme $S \subset X$.

2. $\det \rho = \omega$.

3. ρ is irreducible.

To describe automorphic representations, we need to introduce adèles. If $v \in X$ is a closed point, then the local ring $\mathcal{O}_{X,v}$ is a discrete valuation ring, and determines a valuation $\mathrm{ord}_v : K^\times \to \mathbb{Z}$ which we call a place of K. The completion K_v of K with respect to this valuation is a local field, which can be identified with the field of Laurent series $\mathbb{F}_{q_v}((t_v))$, where $\mathbb{F}_{q_v}$ is the residue field of $\mathcal{O}_{X,v}$ and $t_v \in \mathcal{O}_{X,v}$ is a uniformizing parameter. We write $\mathcal{O}_{K_v} \subset K_v$ for the valuation ring.

The adèle ring $\mathbb{A}_K$ is the restricted direct product of the rings K_v, with respect to their open compact subrings $\mathcal{O}_{K_v}$. It is a locally compact topological ring. Taking adèle points of GL_n, we obtain the group $\mathrm{GL}_n(\mathbb{A}_K)$, which is a locally compact topological group, and which can itself be identified with the restricted direct product of the groups $\mathrm{GL}_n(K_v)$ with respect to their open compact subgroups $\mathrm{GL}_n(\mathcal{O}_{K_v})$.

If $n = 1$, then $\mathrm{GL}_1(\mathbb{A}_K) = \mathbb{A}_K^\times$ and class field theory gives a continuous map

$$\mathrm{Art}_K : K^\times \backslash \mathbb{A}_K^\times \to \Gamma_K^{\mathrm{ab}},$$

which is injective with dense image. We write $\mathcal{A}_{n,\omega}$ for the $\overline{\mathbb{Q}}_\ell$-vector space of functions $f : \mathrm{GL}_n(\mathbb{A}_K) \to \overline{\mathbb{Q}}_\ell$ satisfying the following conditions:

1. f is invariant under left translation by the discrete subgroup $\mathrm{GL}_n(K) \subset \mathrm{GL}_n(\mathbb{A}_K)$.
2. For any $z \in \mathbb{A}_K^\times$, $g \in \mathrm{GL}_n(\mathbb{A}_K)$, $f(gz) = \omega(\mathrm{Art}_K(z))f(g)$.
3. f is smooth, i.e. there exists an open compact subgroup $U \subset \mathrm{GL}_n(\mathbb{A}_K)$ such that for all $u \in U$, $g \in \mathrm{GL}_n(\mathbb{A}_K)$, $f(gu) = f(g)$.

Then the group $\mathrm{GL}_n(\mathbb{A}_K)$ acts on $\mathcal{A}_{n,\omega}$ by right translation. We write $\mathcal{A}_{n,\omega,\mathrm{cusp}} \subset \mathcal{A}_{n,\omega}$ for the subspace of cuspidal functions, i.e. those satisfying the following additional condition:

4. For each proper parabolic subgroup $P \subset \mathrm{GL}_n$, of unipotent radical N, we have

$$\int_{n \in N(K)\backslash N(\mathbb{A}_K)} f(ng)\, dn = 0$$

   for all $g \in \mathrm{GL}_n(\mathbb{A}_K)$. (Note that the quotient $N(K)\backslash N(\mathbb{A}_K)$ is compact, so the integral, taken with respect to a quotient Haar measure on $N(\mathbb{A}_K)$, is well-defined.)

With this definition, $\mathcal{A}_{n,\omega,\mathrm{cusp}} \subset \mathcal{A}_{n,\omega}$ is an $\overline{\mathbb{Q}}_\ell[\mathrm{GL}_n(\mathbb{A}_K)]$-submodule. The following theorem describes the basic structure of this representation of $\mathrm{GL}_n(\mathbb{A}_K)$.

Theorem 2.1. *1. $\mathcal{A}_{n,\omega,cusp}$ is a semisimple admissible $\overline{\mathbb{Q}}_\ell[\mathrm{GL}_n(\mathbb{A}_K)]$-module. Each irreducible constituent $\pi \subset \mathcal{A}_{n,\omega,cusp}$ appears with multiplicity 1.*

2. If $\pi \subset \mathcal{A}_{n,\omega,cusp}$ is an irreducible submodule, then there is a decomposition $\pi = \otimes'_v \pi_v$ of π as a restricted tensor product of irreducible admissible representations π_v of the groups $\mathrm{GL}_n(K_v)$ (where v runs over the set of all places of K).

If π is an irreducible constituent of $\mathcal{A}_{n,\omega,\mathrm{cusp}}$, then we call π a cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_K)$. We write $\mathrm{Aut}_{n,\omega}$ for the set of isomorphism classes of cuspidal automorphic representations of $\mathrm{GL}_n(\mathbb{A}_K)$.

We can now state Langlands reciprocity for GL_n.

Theorem 2.2. *There is a bijection $\mathrm{Gal}_{n,\omega} \leftrightarrow \mathrm{Aut}_{n,\omega}$.*

In order for this theorem to have content, we need to describe how to characterize the bijection whose existence it asserts. The most basic characterization uses restriction to unramified places. Let v be a place of K. If $\rho \in \mathrm{Gal}_{n,\omega}$, then we can consider its restriction $\rho_v = \rho|_{W_{K_v}}$ to the Weil group $W_{K_v} \subset \Gamma_{K_v}$.[2] If $\pi \in \mathrm{Aut}_{n,\omega}$, then we can consider the factor π_v, which is an irreducible admissible representation of the group $\mathrm{GL}_n(K_v)$.

Definition 2.3. *1. A continuous homomorphism $\rho_v : W_{K_v} \to \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$ is said to be unramified if it factors through the unramified quotient $W_{K_v} \to \mathbb{Z}$.*

2. An irreducible admissible representation of the group $\mathrm{GL}_n(K_v)$ is said to be unramified if the subspace $\pi_v^{\mathrm{GL}_n(\mathcal{O}_{K_v})}$ of $\mathrm{GL}_n(\mathcal{O}_{K_v})$-invariant vectors is non-zero.

These two kinds of unramified objects are related by the *unramified local Langlands correspondence*, which can be phrased as follows:

Theorem 2.4. *Let v be a place of K. There is a canonical[3] bijection $\pi_v \mapsto t(\pi_v)$ between the following two sets:*

1. The set of isomorphism classes of unramified irreducible admissible representations π_v of $\mathrm{GL}_n(K_v)$ over $\overline{\mathbb{Q}}_\ell$.

2. The set of semisimple conjugacy classes t in $\mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$.

Proof (sketch). The proof, which is valid for any reductive group G over $\mathbb{F}_q$, goes via the Satake isomorphism. If G is split then this is an isomorphism

$$\mathcal{H}(G(K_v), G(\mathcal{O}_{K_v})) \otimes_{\mathbb{Z}} \mathbb{Z}[q_v^{\pm\frac{1}{2}}] \to \mathbb{Z}[\widehat{G}]^{\widehat{G}} \otimes \mathbb{Z}[q_v^{\pm\frac{1}{2}}],$$

$$T_{v,f} \mapsto f$$

where $\mathcal{H}$ is the Hecke algebra of $G(\mathcal{O}_{K_v})$-biinvariant functions $f : G(K_v) \to \mathbb{Z}$ of compact support, and $\mathbb{Z}[\widehat{G}]^{\widehat{G}}$ is the algebra of conjugation-invariant functions on the dual group $\widehat{G}$ (which is GL_n if $G = \mathrm{GL}_n$). If π_v is an irreducible admissible representation of $G(K_v)$ over $\overline{\mathbb{Q}}_l$ and $\pi_v^{G(\mathcal{O}_{K_v})} \neq 0$, then $\pi_v^{G(\mathcal{O}_{K_v})}$ is a simple $\mathcal{H} \otimes_{\mathbb{Z}} \overline{\mathbb{Q}}_l$-module, which therefore determines a homomorphism $\mathbb{Z}[\widehat{G}]^{\widehat{G}} \to \overline{\mathbb{Q}}_\ell$. The geometric invariant theory of the adjoint quotient of the reductive group $\widehat{G}$ implies that giving such a homomorphism is equivalent to giving a conjugacy class of semisimple elements in $\widehat{G}(\overline{\mathbb{Q}}_\ell)$. □

[2]Here $\Gamma_{K_v} = \mathrm{Gal}(K_v^s/K_v)$ is the absolute Galois group of K_v, with respect to a fixed choice of separable closure. An embedding $K^s \to K_v^s$ extending the map $K \to K_v$ determines an embedding $\Gamma_{K_v} \to \Gamma_K$. The Weil group $W_{K_v} \subset \Gamma_{K_v}$ is the subgroup of elements which act on the residue field by an integral power of the (geometric) Frobenius element Frob_v; see for example Tate [1979] for a detailed discussion.

[3]As the proof shows, we need to fix as well a choice of a square root of q in $\overline{\mathbb{Q}}_\ell$.

Let S be a finite set of places of K. We write $\mathrm{Gal}_{n,\omega,S} \subset \mathrm{Gal}_{n,\omega}$ for the set of ρ such that for each place $v \notin S$ of K, $\rho|_{W_{K_v}}$ is unramified (we say that 'ρ is unramified outside S'). We write $\mathrm{Aut}_{n,\omega,S} \subset \mathrm{Aut}_{n,\omega}$ for the set of $\pi = \otimes'_v \pi_v$ such that for each place $v \notin S$ of K, π_v is unramified (we say that 'π is unramified outside S'). We can now state a more precise version of Theorem 2.2:

Theorem 2.5. *Let S be a finite set of places of K. Then there is a bijection $\pi \mapsto \rho(\pi) : \mathrm{Aut}_{n,\omega,S} \to \mathrm{Gal}_{n,\omega,S}$ with the following property: for each place $v \notin S$, $\rho(\pi)(\mathrm{Frob}_v)^{ss} \in t(\pi_v)$*[4].

This defining property uniquely characterizes the bijection, if it exists. Indeed, the isomorphism class of any representation $\pi \in \mathrm{Aut}_{n,\omega,S}$ is uniquely determined by the representations π_v $(v \notin S)$: this is the strong multiplicity one theorem. Similarly, any representation $\rho \in \mathrm{Gal}_{n,\omega,S}$ is uniquely determined by the conjugacy classes of the elements $\rho(\mathrm{Frob}_v)^{\mathrm{ss}}$ $(v \notin S)$: the irreducible representation ρ is uniquely determined up to isomorphism by its character $\mathrm{tr}\,\rho$. This continuous function $\mathrm{tr}\,\rho : \Gamma_{K,S} \to \overline{\mathbb{Q}}_\ell$ is determined by its values at a dense set of elements, and the Chebotarev density theorem implies that the Frobenius elements Frob_v $(v \notin S)$ form such a set.

L. Lafforgue proved Theorem 2.5 using an induction on n. If the theorem is known for $n' < n$, then the 'principe de récurrence de Deligne' (see L. Lafforgue [2002, Appendice B]) reduces the problem to constructing, for any $\pi \in \mathrm{Aut}_{n,\omega,S}$, the corresponding Galois representation $\rho(\pi) \in \mathrm{Gal}_{n,\omega,S}$, as well as proving that certain L– and ϵ–factors are matched up under the correspondence. The three main ingredients that make this possible are Grothendieck's theory of L-functions of Galois representations, Laumon's product formula for the ϵ–factors of Galois representations, and Piatetski-Shapiro's converse theorem, which can be used to show that an irreducible admissible representation of $\mathrm{GL}_n(\mathbb{A}_K)$ with sufficiently well-behaved associated L-functions is in fact cuspidal automorphic. We note that in carrying this out Lafforgue actually obtains a much more precise result than Theorem 2.5, in particular re-proving the local Langlands correspondence for GL_n and showing that the global correspondence is compatible with the local one.

One would like to generalise Theorem 2.5 to an arbitrary reductive group G over $\mathbb{F}_q$. However, there are a number (!) of difficulties. To begin with, it is not even clear what the correct statement should be: it is easy to write down the naive analogues of the sets $\mathrm{Aut}_{n,\omega}$ and $\mathrm{Gal}_{n,\omega}$, but we will see some reasons why they cannot be related by a simple bijection. Moreover, no converse theorem is known for a general group G, which means there is no apparent way of proving that a given admissible representation of $G(\mathbb{A}_K)$ is in fact cuspidal automorphic. This is the motivation behind proving a result like our main theorem.

[4]Here and elsewhere, we write x^{ss} for the semisimple part in the Jordan decomposition $x = x^{\mathrm{ss}}x^{\mathrm{u}}$ of an element x of a linear algebraic group.

3 Pseudocharacters

Let us now pass to the case of a general reductive group G over $\mathbb{F}_q$. In order to simplify the discussion here, we will assume that G is split. In this case one can associate to G its dual group $\widehat{G}$, a split reductive group over $\mathbb{Z}$, which is characterized by the property that its root datum is dual to that of G (see e.g. Borel [1979]). If $G = \mathrm{GL}_n$, then $\widehat{G} = \mathrm{GL}_n$, so our discussion will include n-dimensional linear representations as a special case.

We will describe the results of Vincent Lafforgue's construction in the next section. First, we make a detour to describe the notion of $\widehat{G}$-pseudocharacter, which was introduced for the first time in V. Lafforgue [n.d.]. This is a generalization of the notion of the pseudocharacter of an n-dimensional representation (to which it reduces in the case $G = \mathrm{GL}_n$).

Let Γ be a group, and let Ω be an algebraically closed field of characteristic 0. We recall that to any representation $\rho : \Gamma \to \mathrm{GL}_n(\Omega)$, we can associate the character $\operatorname{tr} \rho : \Gamma \to \Omega$; it clearly depends only on ρ up to conjugacy and up to semisimplification. We have the following theorem, the second part of which was proved by Taylor using results of Taylor [1991] and Procesi [1976].

Theorem 3.1. *1. Let $\rho, \rho' : \Gamma \to \mathrm{GL}_n(\Omega)$ be semisimple representations. Then they are isomorphic if and only if* $\operatorname{tr} \rho = \operatorname{tr} \rho'$.

2. Let $t : \Gamma \to \Omega$ be a function satisfying the following conditions:

(a) $t(1) = n$.

(b) For all $\gamma_1, \gamma_2 \in \Gamma$, $t(\gamma_1\gamma_2) = t(\gamma_2\gamma_1)$.

(c) For all $\gamma_1, \dots, \gamma_{n+1} \in \Gamma$, $\sum_{\sigma \in S_{n+1}} t_\sigma(\gamma_1, \dots, \gamma_{n+1}) = 0$, where if σ has cycle decomposition

$$\sigma = (a_1 \dots a_{k_1})(b_1 \dots b_{k_2}) \dots$$

then we set

$$t_\sigma(\gamma_1, \dots, \gamma_{n+1}) = t(\gamma_{a_1} \dots \gamma_{a_{k_1}}) t(\gamma_{b_1} \dots \gamma_{b_{k_2}}) \dots .$$

Then there exists a representation $\rho : \Gamma \to \mathrm{GL}_n(\Omega)$ such that $\operatorname{tr} \rho = t$.

We can call a function $t : \Gamma \to \Omega$ satisfying the condition of Theorem 3.1 a pseudocharacter of dimension n. Then the theorem says that sets of conjugacy classes of semisimple representations $\rho : \Gamma \to \mathrm{GL}_n(\Omega)$ and of pseudocharacters of dimension n are in canonical bijection.

Here is Lafforgue's definition of a $\widehat{G}$-pseudocharacter. Let A be a ring.

Definition 3.2. *Let* $\mathbf{t} = (t_n)_{n \geq 1}$ *be a collection of algebra maps* $t_n : \mathbb{Z}[\widehat{G}^n]^{\widehat{G}} \to \mathrm{Fun}(\Gamma^n, A)$ *satisfying the following conditions:*

1. *For each* $n, m \geq 1$ *and for each* $\zeta : \{1, \dots, m\} \to \{1, \dots, n\}$, $f \in \mathbb{Z}[\widehat{G}^m]^{\widehat{G}}$, *and* $\gamma = (\gamma_1, \dots, \gamma_n) \in \Gamma^n$, *we have*
$$t_n(f^\zeta)(\gamma_1, \dots, \gamma_n) = t_m(f)(\gamma_{\zeta(1)}, \dots, \gamma_{\zeta(m)}),$$
where $f^\zeta(g_1, \dots, g_n) = f(g_{\zeta(1)}, \dots, g_{\zeta(m)})$.

2. *For each* $n \geq 1$, $\gamma = (\gamma_1, \dots, \gamma_{n+1}) \in \Gamma^{n+1}$, *and* $f \in \mathbb{Z}[\widehat{G}^n]^{\widehat{G}}$, *we have*
$$t_{n+1}(\hat{f})(\gamma_1, \dots, \gamma_{n+1}) = t_n(f)(\gamma_1, \dots, \gamma_{n-1}, \gamma_n\gamma_{n+1}),$$
where $\hat{f}(g_1, \dots, g_{n+1}) = f(g_1, \dots, g_{n-1}, g_n g_{n+1})$.

Then $\mathbf{t}$ *is called a* $\widehat{G}$*-pseudocharacter of* Γ *over* A.

Note that $\widehat{G}$ acts on $\widehat{G}^n$ by diagonal conjugation. The subring $\mathbb{Z}[\widehat{G}^n]^{\widehat{G}} \subset \mathbb{Z}[\widehat{G}^n]$ is the ring of functions invariant under this action. We observe that if $\rho : \Gamma \to \widehat{G}(A)$ is a homomorphism, then we can define a $\widehat{G}$-pseudocharacter $\operatorname{tr} \rho = (t_n)_{n \geq 1}$ of Γ over A by the formula
$$t_n(f)(\gamma_1, \dots, \gamma_n) = f(\rho(\gamma_1), \dots, \rho(\gamma_n)).$$
It is clear that this depends only on the $\widehat{G}(A)$-conjugacy class of ρ.

Theorem 3.3. *Let* Γ *be a group, and let* Ω *be an algebraically closed field.*

1. *Let* $\rho, \rho' : \Gamma \to \widehat{G}(\Omega)$ *be* $\widehat{G}$*-completely reducible representations.*[5] *Then* ρ, ρ' *are* $\widehat{G}(\Omega)$*-conjugate if and only if* $\operatorname{tr} \rho = \operatorname{tr} \rho'$.

2. *Let* $\mathbf{t}$ *be a* $\widehat{G}$*-pseudocharacter. Then there exists a representation* $\rho : \Gamma \to \widehat{G}(\Omega)$ *such that* $\mathbf{t} = \operatorname{tr} \rho$.

The proof of Theorem 3.3 is based on Richardson's results about the geometric invariant theory of the action of $\widehat{G}$ on $\widehat{G}^n$ by diagonal conjugation Richardson [1988].

In the case where Γ is profinite, we want to impose continuity conditions on its $\widehat{G}$-pseudocharacters. Fortunately, $\widehat{G}$-pseudocharacters are well-behaved from this point of view.

[5] A representation ρ is said to be $\widehat{G}$-irreducible if its image is contained in no proper parabolic subgroup of $\widehat{G}_\Omega$, and $\widehat{G}$-completely reducible if for any parabolic subgroup $P \subset \widehat{G}_\Omega$ containing the image, there exists a Levi subgroup $L \subset P$ such that $\rho(\Gamma) \subset L(\Omega)$. See e.g. Serre [2005].

Definition 3.4. *Let A be a topological ring, and let Γ be a profinite group. We say that a $\widehat{G}$-pseudocharacter $\mathbf{t}$ of Γ over A is continuous if each map $t_n : \mathbb{Z}[\widehat{G}^n]^{\widehat{G}} \to \mathrm{Fun}(\Gamma^n, A)$ takes values in the subset of continuous functions $\mathrm{Fun}_{\mathrm{cts}}(\Gamma^n, A)$.*

Proposition 3.5. *Let ℓ be a prime, and let Γ be a profinite group. Let $\Omega = \overline{\mathbb{Q}}_\ell$ (with its ℓ-adic topology) or $\overline{\mathbb{F}}_\ell$ (with the discrete topology). Let $\rho : \Gamma \to \widehat{G}(\Omega)$ be a $\widehat{G}$-completely reducible representation. Then ρ is continuous if and only if $\operatorname{tr} \rho$ is continuous.*

Finally, we note that $\widehat{G}$-pseudocharacters are well-behaved from the point of view of reduction modulo ℓ. We will need this in our discussion of the deformation theory of pseudocharacters later on.

Proposition 3.6. *Let ℓ be a prime, and let Γ be a profinite group. Then:*

1. *Let $\mathbf{t}$ be a continuous $\widehat{G}$-pseudocharacter of Γ over $\overline{\mathbb{Q}}_\ell$. Then $\mathbf{t}$ takes values in $\overline{\mathbb{Z}}_\ell$ and $\bar{\mathbf{t}}$, its reduction modulo ℓ, is a continuous $\widehat{G}$-pseudocharacter over $\overline{\mathbb{F}}_\ell$.*

2. *Let $\rho : \Gamma \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a continuous representation. After replacing ρ by a $\widehat{G}(\overline{\mathbb{Q}}_\ell)$-conjugate, we can assume that ρ takes values in $\widehat{G}(\overline{\mathbb{Z}}_\ell)$. Let $\overline{\rho} : \Gamma \to \widehat{G}(\overline{\mathbb{F}}_\ell)$ denote the semisimplification of the reduction of ρ modulo ℓ. Then $\overline{\rho}$ depends only on ρ up to $\widehat{G}(\overline{\mathbb{Q}}_\ell)$-conjugacy, and $\operatorname{tr} \overline{\rho} = \overline{\operatorname{tr} \rho}$.*

We now come back to our original case of interest, namely pseudocharacters of the group $\Gamma = \Gamma_{K,S}$. We can define a compatible family of pseudocharacters of $\Gamma_{K,S}$ of dimension n to consist of the data of a number field E and, for each prime-to-q place λ of E, a pseudocharacter $t_\lambda : \Gamma_{K,S} \to E_\lambda$ of dimension n. These should satisfy the following property:

- For each place $v \notin S$ of K, the number $t_\lambda(\mathrm{Frob}_v)$ lies in $E \subset E_\lambda$ and is independent of the choice of λ.

Thus, for example, Deligne [1980, Conjecture I.2.10] asks that every pseudocharacter $t_\ell : \Gamma_{K,S} \to \overline{\mathbb{Q}}_\ell$ satisfying certain conditions should be a member of a compatible family. This leads us to our first question:

Question 3.7. *Is it possible to define a notion of 'compatible family of $\widehat{G}$-pseudocharacters', generalizing the above notion for GL_n?*

If we are willing to consider instead compatible families of representations, then Drinfeld [n.d.] gives a satisfying (positive) answer to Question 3.7 using the results of L. Lafforgue [2002]. The question remains, however, of whether we can phrase this for pseudocharacters in elementary terms and, in particular, whether it is possible to make sense of compatible families when K is instead a global field of characteristic 0 (i.e. a number field).

Here is one case it is easy to make sense of the notion of compatible family:

Proposition 3.8. *Let $\ell \nmid q$ be a prime, and let $\rho : \Gamma_{K,S} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a continuous representation of Zariski dense image. Assume that $\widehat{G}$ is semisimple. Then we can find a number field E and an embedding $E \hookrightarrow \overline{\mathbb{Q}}_\ell$, inducing the place λ_0 of E, with the following properties:*

- *For each place $v \notin S$ of K and for each $f \in \mathbb{Z}[\widehat{G}]^{\widehat{G}}$, $f(\rho(\mathrm{Frob}_v)) \in E$. In other words, the conjugacy class of $\rho(\mathrm{Frob}_v)^{ss}$ is defined over E.*
- *For each prime-to-q place λ of E, there exists a continuous homomorphism $\rho_\lambda : \Gamma_{K,S} \to \widehat{G}(E_\lambda)$ of Zariski dense image such that for each place $v \notin S$ of K and for each $f \in \mathbb{Z}[\widehat{G}]^{\widehat{G}}$, $f(\rho_\lambda(\mathrm{Frob}_v))$ lies in E and equals $f(\rho(\mathrm{Frob}_v))$. In other words, $\rho_\lambda(\mathrm{Frob}_v)^{ss}$ lies in the same geometric conjugacy class as $\rho(\mathrm{Frob}_v)^{ss}$.*

Furthermore, for any prime-to-q place λ of E and for any continuous homomorphism $\rho'_\lambda : \Gamma_{K,S} \to \widehat{G}(\overline{E}_\lambda)$ such that for each place $v \notin S$ of K and for each $f \in \mathbb{Z}[\widehat{G}]^{\widehat{G}}$, $f(\rho'_\lambda(\mathrm{Frob}_v))$ lies in E and equals $f(\rho(\mathrm{Frob}_v))$, ρ'_λ is $\widehat{G}(\overline{E}_\lambda)$-conjugate to ρ_λ. In particular, ρ and ρ_{λ_0} are $\widehat{G}(\overline{\mathbb{Q}}_\ell)$-conjugate.

The proof makes use of the proof of the global Langlands correspondence for GL_n by L. Lafforgue [ibid.], together with Chin's application of this work to the analysis of compatible families Chin [2004]; see Böckle, Harris, Khare, and J. A. Thorne [n.d., §6].

4 The work of V. Lafforgue

Having introduced the notion of $\widehat{G}$-pseudocharacter, we can now describe the basic shape of Vincent Lafforgue's results in V. Lafforgue [n.d.]. We recall that G is a split reductive group over $\mathbb{F}_q$. In order to simplify statements, we are now going to impose the further assumption that G has finite centre (i.e. is semisimple).[6] We write $\mathfrak{A}_G$ for the $\overline{\mathbb{Q}}_\ell$-vector space of functions $f : G(\mathbb{A}_K) \to \overline{\mathbb{Q}}_\ell$ satisfying the following conditions:

1. f is invariant under left translation by the discrete subgroup $G(K) \subset G(\mathbb{A}_K)$.
2. f is smooth.

[6] By contrast, the paper V. Lafforgue [n.d.] does not impose any restriction on G; see in particular §12 of *op. cit.*

Then the group $G(\mathbb{A}_K)$ acts on $\mathcal{A}_G$ by right translation. We write $\mathcal{A}_{G,\text{cusp}} \subset \mathcal{A}_G$ for the subspace of cuspidal functions, i.e. those satisfying the following condition:

3. For any proper parabolic subgroup $P \subset G$ of unipotent radical N, we have
$$\int_{n \in N(K)\backslash N(\mathbb{A}_K)} f(ng)\, dn = 0$$
for all $g \in G(\mathbb{A}_K)$.

With this definition, $\mathcal{A}_{G,\text{cusp}}$ is a semisimple admissible $\overline{\mathbb{Q}}_\ell[G(\mathbb{A}_K)]$-module. In general, understanding the constituents of this space is much more complicated than for the group GL_n. For example:

- Multiplicity one does not hold: there can exist representations π of $G(\mathbb{A}_K)$ which appear in $\mathcal{A}_{G,\text{cusp}}$ with multiplicity greater than 1.
- Strong multiplicity one does not hold: there can exist representation π, π' of $G(\mathbb{A}_K)$ which have positive multiplicity in $\mathcal{A}_{G,\text{cusp}}$, such that $\pi_v \cong \pi'_v$ for all but finitely many places v of K, but such that $\pi \not\cong \pi'$.

These phenomena are reflected in what happens on the Galois side. For example:

- There can exist everywhere unramified homomorphisms $\rho, \rho' : \Gamma_K \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ such that $\rho(\mathrm{Frob}_v)$ and $\rho'(\mathrm{Frob}_v)$ are conjugate for every v, but such that ρ, ρ' are not conjugate.
- There can exist homomorphisms $\rho, \rho' : \Gamma_K \to \widehat{G}$ and a place v_0 of K such that for all $v \neq v_0$, ρ and ρ' are unramified at v and $\rho(\mathrm{Frob}_v)$, $\rho'(\mathrm{Frob}_v)$ are $\widehat{G}(\overline{\mathbb{Q}}_\ell)$-conjugate; but $\rho|_{\Gamma_{K_{v_0}}} \not\cong \rho'|_{\Gamma_{K_{v_0}}}$. (Another reason for the failure of strong multiplicity one, not related to this Galois–theoretic phenomenon, is the existence of non-trivial L-packets.)

We refer the reader to Wang [2012] for a survey of how the relation between these phenomena can be understood in terms of Arthur's conjectural decomposition of the space of automorphic forms in terms of A-parameters Arthur [1989].

Lafforgue's construction, quite remarkably, gives a decomposition of the space $\mathcal{A}_{G,\text{cusp}}$ of cusp forms on G which is quite close in appearance to that predicted by Arthur. He defines for each $n \geq 1$, function $f \in \mathbb{Z}[\widehat{G}^n]^{\widehat{G}}$, and tuple of elements $\gamma = (\gamma_1, \dots, \gamma_n) \in \Gamma_K^n$, an operator $S_{n,f,\gamma} \in \mathrm{End}_{\overline{\mathbb{Q}}_\ell}(\mathcal{A}_{G,\text{cusp}})$. He calls these 'excursion operators', and proves the following two theorems:

Theorem 4.1. *1. The operators $S_{n,f,\gamma}$ commute with each other and with the action of $G(\mathbb{A}_K)$.*

2. *Let* $\mathcal{B} \subset \mathrm{End}_{\overline{\mathbb{Q}}_\ell}(\mathcal{A}_{G,cusp})$ *denote the* $\overline{\mathbb{Q}}_\ell$*-subalgebra generated by the operators* $S_{n,f,\gamma}$ *for all possible choices of* n, f, *and* γ. *Then the system of maps* $\mathbf{t} = (t_n)_{n \geq 1}$ *given by*

$$t_n : \mathbb{Z}[\widehat{G}^n]^{\widehat{G}} \to \mathrm{Fun}(\Gamma_K^n, \mathcal{B}),$$
$$f \mapsto (\gamma \mapsto S_{n,f,\gamma})$$

is a $\widehat{G}$*-pseudocharacter of* Γ_K *valued in* $\mathcal{B}$.

Theorem 4.2. *Let* S *be a finite set of places of* K, *and let* $U \subset G(\mathbb{A}_K)$ *be an open compact subgroup such that for each place* $v \notin S$ *of* K, $U_v = G(\mathcal{O}_{K_v})$. *Let* $\mathcal{B}_U$ *denote the quotient of* $\mathcal{B}$ *which acts faithfully on* $\mathcal{A}_{G,cusp,U}$. *Then:*

1. *The pushforward of* $\mathbf{t}$ *along* $\mathcal{B} \to \mathcal{B}_U$ *is pulled back from a* $\widehat{G}$*-pseudocharacter* $\mathbf{t}_U$ *of* $\Gamma_{K,S}$ *valued in* $\mathcal{B}_U$.

2. *If* $v \notin S$, *then the image of* S_{1,f,Frob_v} *in* $\mathcal{B}_U$ *equals the unramified Hecke operator* $T_{v,f}$ *(defined as in the proof of Theorem 2.4, via the Satake isomorphism).*

Since the algebra $\mathcal{B}_U$ contains the unramified Hecke operators, it can be viewed as an enlargement of the usual Hecke algebra.

Corollary 4.3. *Let* π *be a cuspidal automorphic representation of* $G(\mathbb{A}_K)$, *and let* $V_\pi \subset \mathcal{A}_{G,cusp}$ *be the* π*-isotypic component. Let* $\mathcal{B}_\pi$ *denote the quotient of* $\mathcal{B}$ *which acts faithfully on* V_π. *Then* $\mathcal{B}_\pi$ *is a finite-dimensional* $\overline{\mathbb{Q}}_\ell$*-algebra and for each maximal ideal* $\mathfrak{p}$, *one can associate a continuous representation* $\sigma_{\pi,\mathfrak{p}} : \Gamma_K \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ *with the following properties:*

1. $\mathrm{tr}\,\sigma_{\pi,\mathfrak{p}} = \mathbf{t}_\pi$ *mod* $\mathfrak{p}$.

2. *Let* S *be a finite set of places of* K *such that* $\pi_v^{G(\mathcal{O}_{K_v})} \neq 0$ *if* $v \notin S$. *Then* $\sigma_{\pi,\mathfrak{p}}$ *is unramified outside* S *and if* $v \notin S$, *then* $\sigma_{\pi,\mathfrak{p}}(\mathrm{Frob}_v)^{ss} \in t(\pi_v)$.

3. *If* $\mathfrak{p} \neq \mathfrak{p}'$, *then* $\sigma_{\pi,\mathfrak{p}} \not\cong \sigma_{\pi,\mathfrak{p}'}$.

Proof. Since V_π has finite length as a $\overline{\mathbb{Q}}_\ell[G(\mathbb{A}_K)]$-module and $\mathcal{B}_\pi$ is contained inside $\mathrm{End}_{\overline{\mathbb{Q}}_\ell[G(\mathbb{A}_K)]}(V_\pi)$, $\mathcal{B}_\pi$ is a finite-dimensional $\overline{\mathbb{Q}}_\ell$-algebra. Since $\mathcal{B}_\pi$ is a quotient of $\mathcal{B}$, it carries a $\widehat{G}$-pseudocharacter $\mathbf{t}_\pi$. Each maximal ideal $\mathfrak{p} \subset \mathcal{B}_\pi$ has residue field $\overline{\mathbb{Q}}_\ell$, and the pushforward of $\mathbf{t}_\pi$ along the map $\mathcal{B}_\pi \to \mathcal{B}_\pi/\mathfrak{p} \cong \overline{\mathbb{Q}}_\ell$ therefore corresponds, by Theorem 3.3, to a continuous $\widehat{G}$-completely reducible representation $\sigma_{\pi,\mathfrak{p}} : \Gamma_{K,S} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ satisfying the following property: for all $n \geq 1$, $f \in \mathbb{Z}[\widehat{G}^n]^{\widehat{G}}$, $\gamma = (\gamma_1, \ldots, \gamma_n) \in \Gamma_{K,S}^n$, we have

$$f(\sigma_{\pi,\mathfrak{p}}(\gamma_1), \ldots, \sigma_{\pi,\mathfrak{p}}(\gamma_n)) = S_{n,f,\gamma} \bmod \mathfrak{p}.$$

From this identity it is apparent that $\sigma_{\pi,\mathfrak{p}}$ determines $\mathfrak{p}$. Specializing to $n = 1$ and $\gamma = \mathrm{Frob}_v$ for some $v \notin S$, this identity reduces to the formula

$$f(\sigma_{\pi,\mathfrak{p}}(\mathrm{Frob}_v)) = T_{v,f} \bmod \mathfrak{p},$$

or equivalently that $\sigma_{\pi,\mathfrak{p}}(\mathrm{Frob}_v)^{\mathrm{ss}}$ is in the conjugacy class $t(\pi_v)$. □

Question 4.4. *The space V_π can be defined over $\overline{\mathbb{Q}}$. Is there a sense in which its decomposition $V_\pi = \oplus V_{\pi,\mathfrak{p}}$ is independent of ℓ?*

Presumably a positive answer to this question must be tied up with a positive answer to Question 3.7.

We can now define what it means for a Galois representation to be cuspidal automorphic, in the sense of the algebra $\mathfrak{B}$.

Definition 4.5. *We say that a representation $\rho : \Gamma_K \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ is cuspidal automorphic if there exists a cuspidal automorphic representation π of $G(\mathbb{A}_K)$ and a maximal ideal $\mathfrak{p} \subset \mathfrak{B}_\pi$ such that $\rho \cong \sigma_{\pi,\mathfrak{p}}$.*

Note that this definition depends in an essential way on Lafforgue's excursion operators!

We are now in a position to state the main theorem of Böckle, Harris, Khare, and J. A. Thorne [n.d.]:

Theorem 4.6. *Let $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a continuous representation of Zariski dense image. Then ρ is* potentially cuspidal automorphic*: there exists a finite Galois extension L/K such that $\rho|_{\Gamma_{L,\varnothing}}$ is cuspidal automorphic in the sense of Definition 4.5.*

Corollary 4.7. *Let $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a continuous representation of Zariski dense image. Then there exists a finite Galois extension L/K and an everywhere unramified cuspidal automorphic representation π of $G(\mathbb{A}_L)$ such that for each place w of L, $\rho|_{W_{L_w}}$ and π_w are related under the unramified local Langlands correspondence for $G(L_w)$.*

(In fact, Proposition 3.8 implies that the theorem and its corollary are equivalent.) In the remainder of this article we will sketch the proof of Theorem 4.6.

5 An automorphy lifting theorem for G

How can one show that a Galois representation $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ is automorphic, in the sense of Definition 4.5? For a general group G, we no longer know how to construct automorphic forms using converse theorems.

We pursue a different path which is inspired by the proofs of existing potential automorphy results for Galois representations $\Gamma_E \to \mathrm{GL}_n$, where E is a number field. These are in turn based on automorphy lifting theorems, which are provable instances of the following general principle:

Principle 5.1. *Let $\rho, \rho' : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be continuous representations and let $\overline{\rho}$, $\overline{\rho}' : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{F}}_\ell)$ denote their reductions modulo ℓ. Suppose that $\overline{\rho}$, $\overline{\rho}'$ are $\widehat{G}(\overline{\mathbb{F}}_\ell)$-conjugate and $\widehat{G}$-irreducible. Suppose that ρ is cuspidal automorphic. Then ρ' is also cuspidal automorphic.*

The first theorem of this type was stated by Wiles on his way to proving Fermat's Last Theorem Wiles [1995]. Our proof of an analogous result is inspired by Diamond's elaboration of the Taylor–Wiles method Diamond [1997], which gives a way to construct an isomorphism $R \cong \mathbb{T}$, where R is a Galois deformation ring and $\mathbb{T}$ is a Hecke algebra acting on cuspidal automorphic forms. By contrast, we prove an '$R = \mathfrak{B}$' theorem, where $\mathfrak{B}$ is a suitable ring of Lafforgue's excursion operators.

We describe these objects in order to be able to state a precise result. We will stick to the everywhere unramified case. We first consider the Galois side. Let $k \subset \overline{\mathbb{F}}_l$ be a finite subfield, and let $\overline{\rho} : \Gamma_{K,\varnothing} \to \widehat{G}(k)$ be a continuous homomorphism. Let Art_k denote the category of Artinian local $W(k)$-algebras A, equipped with an isomorphism $A/\mathfrak{m}_A \cong k$. We define $\mathrm{Lift}_{\overline{\rho}} : \mathrm{Art}_k \to \mathrm{Sets}$ to be the functor of liftings of $\overline{\rho}$, i.e. of homomorphisms $\rho_A : \Gamma_{K,\varnothing} \to \widehat{G}(A)$ such that $\rho_A \bmod \mathfrak{m}_A = \overline{\rho}$.

For any $A \in \mathrm{Art}_k$, the group $\ker(\widehat{G}(A) \to \widehat{G}(k))$ acts on $\mathrm{Lift}_{\overline{\rho}}(A)$ by conjugation, and we write $\mathrm{Def}_{\overline{\rho}} : \mathrm{Art}_k \to \mathrm{Sets}$ for the quotient functor (given by the formula $\mathrm{Def}_{\overline{\rho}}(A) = \mathrm{Lift}_{\overline{\rho}}(A)/\ker(\widehat{G}(A) \to \widehat{G}(k))$). The following lemma is basic.

Lemma 5.2. *Suppose that $\overline{\rho}$ is absolutely $\widehat{G}$-irreducible, and that ℓ does not divide the order of the Weyl group of $\widehat{G}$. Then the functor $\mathrm{Def}_{\overline{\rho}}$ is pro-represented by a complete Noetherian local $W(k)$-algebra $R_{\overline{\rho}}$ with residue field k.*

In order to be able to relate the deformation ring $R_{\overline{\rho}}$ to automorphic forms, we need to introduce integral structures. We therefore write $\mathcal{C}_{G,k}$ for the set of functions $f : G(\mathbb{A}_K) \to W(k)$ satisfying the following conditions:

1. f is invariant under left translation by $G(K)$.

2. f is smooth.

We write $\mathcal{C}_{G,k,\mathrm{cusp}}$ for the intersection $\mathcal{C}_{G,k} \cap \mathcal{A}_{G,\mathrm{cusp}}$ (taken inside $\mathcal{A}_G$). Let $U = \prod_v G(\mathcal{O}_{K_v})$.

Proposition 5.3. *Suppose that $f \in \mathbb{Z}[\widehat{G}^n]^{\widehat{G}}$. Then each operator $S_{n,f,\gamma} \in \mathfrak{B}_U \subset \mathrm{End}_{\overline{\mathbb{Q}}_\ell}(\mathcal{A}^U_{G,cusp})$ leaves invariant the submodule $\mathcal{C}^U_{G,k,cusp}$.*

We write $\mathbb{B}(U, W(k))$ for the $W(k)$-subalgebra of $\mathrm{End}_{W(k)}(\mathcal{C}^U_{G,k,\mathrm{cusp}})$ generated by the operators $S_{n,f,\gamma}$ for $f \in \mathbb{Z}[\widehat{G}^n]^{\widehat{G}}$. Then $\mathbb{B}(U, W(k))$ is a finite flat $W(k)$-algebra and there is a $\widehat{G}$-pseudocharacter $\mathbf{t}_{U,W(k)}$ of $\Gamma_{K,\varnothing}$ valued in $\mathbb{B}(U, W(k))$.

Let $\mathfrak{m}$ be a maximal ideal of $\mathbb{B}(U, W(k))$. Its residue field is a finite extension of the finite field k. After possibly enlarging k, we can assume that the following conditions hold:

- The residue field of $\mathfrak{m}$ equals k.
- There exists a continuous representation $\overline{\rho}_{\mathfrak{m}} : \Gamma_{K,\varnothing} \to \widehat{G}(k)$ such that $\mathrm{tr}\, \overline{\rho}_{\mathfrak{m}} = \mathbf{t}_{U,W(k)} \bmod \mathfrak{m}$.

Then the ring $\mathbb{B}(U, W(k))_{\mathfrak{m}}$ (localization at the maximal ideal $\mathfrak{m}$) is a finite flat local $W(k)$-algebra of residue field k, and it comes equipped with a pseudocharacter $\mathbf{t}_{U,W(k),\mathfrak{m}}$. A natural question to ask is: under what conditions does this pseudocharacter arise from a representation $\rho_{\mathfrak{m}} : \Gamma_{K,\varnothing} \to \widehat{G}(\mathbb{B}(U, W(k))_{\mathfrak{m}})$ lifting $\overline{\rho}_{\mathfrak{m}}$? In other words, under what conditions does the analogue of Theorem 3.3 hold when we no longer restrict to field-valued $\widehat{G}$-pseudocharacters?

Proposition 5.4. *Suppose that $\overline{\rho}_{\mathfrak{m}}$ is absolutely $\widehat{G}$-irreducible, and that its centralizer $\mathrm{Cent}(\widehat{G}^{ad}_k, \overline{\rho}_{\mathfrak{m}})$ is scheme-theoretically trivial.*[7] *Suppose that ℓ does not divide the order of the Weyl group of $\widehat{G}$. Then there is a unique conjugacy class of liftings $[\rho_{\mathfrak{m}}] \in \mathrm{Def}_{\overline{\rho}_{\mathfrak{m}}}(\mathbb{B}(U, W(k))_{\mathfrak{m}})$ such that $\mathrm{tr}\, \rho_{\mathfrak{m}} = \mathbf{t}_{U,W(k),\mathfrak{m}}$.*

Under the assumptions of Proposition 5.4, we see the ring $R_{\overline{\rho}_{\mathfrak{m}}}$ is defined, and that its universal property determines a canonical map $R_{\overline{\rho}_{\mathfrak{m}}} \to \mathbb{B}(U, W(k))_{\mathfrak{m}}$. The localized space $(\mathcal{C}^U_{G,k,\mathrm{cusp}})_{\mathfrak{m}}$ of automorphic forms then becomes a module for the deformation ring $R_{\overline{\rho}_{\mathfrak{m}}}$.

We are now in a position to state a provable instance of Principle 5.1.

Theorem 5.5. *Let $\mathfrak{m} \subset \mathbb{B}(U, W(k))$ be a maximal ideal of residue field k, and suppose that there exists a continuous, absolutely $\widehat{G}$-irreducible representation $\overline{\rho}_{\mathfrak{m}} : \Gamma_{K,\varnothing} \to \widehat{G}(k)$ such that $\mathrm{tr}\, \overline{\rho}_{\mathfrak{m}} = \mathbf{t}_{U,W(k)} \bmod \mathfrak{m}$. Suppose further that the following conditions are satisfied:*

1. *$\ell > \#W$, where W is the Weyl group of the split reductive group $\widehat{G}$.*
2. *The centralizer $\mathrm{Cent}(\widehat{G}^{ad}_k, \overline{\rho}_{\mathfrak{m}})$ is scheme-theoretically trivial.*
3. *The representation $\overline{\rho}_{\mathfrak{m}}$ is absolutely strongly $\widehat{G}$-irreducible.*

[7] Here and elsewhere, $\widehat{G}^{\mathrm{ad}}_k$ denotes the adjoint group of $\widehat{G}_k$, i.e. the quotient of $\widehat{G}_k$ by its centre.

4. *The subgroup $\overline{\rho}_{\mathfrak{m}}(\Gamma_{K(\zeta_\ell)}) \subset \widehat{G}(k)$ is $\widehat{G}$-abundant.*

Then $(\mathcal{C}^U_{G,k,cusp})_{\mathfrak{m}}$ is a finite free $R_{\overline{\rho}_{\mathfrak{m}}}$-module.

Corollary 5.6. *With the assumptions of the theorem, let $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a continuous homomorphism such that $\overline{\rho} \cong \overline{\rho}_{\mathfrak{m}}$. Then ρ is cuspidal automorphic.*

Proof. The theorem implies that the map $R_{\overline{\rho}_{\mathfrak{m}}} \to \mathcal{B}(U, W(k))_{\mathfrak{m}}$ that we have constructed is an isomorphism. Any representation ρ as in the statement of the corollary determines a homomorphism $R_{\overline{\rho}_{\mathfrak{m}}} \to \overline{\mathbb{Q}}_\ell$. (To show this, we first need to prove that a conjugate of ρ takes values in $\widehat{G}(\mathcal{O})$, where $\mathcal{O}$ is a complete Noetherian local $W(k)$-subalgebra of $\overline{\mathbb{Q}}_\ell$ of residue field k.) This in turn determines a homomorphism $\mathcal{B}(U, W(k))_{\mathfrak{m}} \to \overline{\mathbb{Q}}_\ell$, hence a maximal ideal $\mathfrak{p} \subset \mathcal{B}_U$ with the property that for each $n \geq 1$, $f \in \mathbb{Z}[\widehat{G}^n]^{\widehat{G}}$ and $\gamma = (\gamma_1, \dots, \gamma_n) \in \Gamma^n_{K,\varnothing}$,

$$f(\rho(\gamma_1), \dots, \rho(\gamma_n)) = S_{n,f,\gamma} \bmod \mathfrak{p}.$$

This is exactly what it means for ρ be cuspidal automorphic. □

There are two adjectives in the theorem that have yet to be defined: 'strongly $\widehat{G}$-irreducible' and '$\widehat{G}$-abundant'. We remedy this now:

Definition 5.7. *Let Ω be an algebraically closed field, and let Γ be a group. We say that a homomorphism $\sigma : \Gamma \to \widehat{G}(\Omega)$ is strongly $\widehat{G}$-irreducible if for any other homomorphism $\sigma' : \Gamma \to \widehat{G}(\Omega)$ such that for all $\gamma \in \Gamma$, $\sigma(\gamma)^{ss}$ and $\sigma'(\gamma)^{ss}$ are $\widehat{G}(\Omega)$-conjugate, σ' is $\widehat{G}$-irreducible.*

Thus a strongly $\widehat{G}$-irreducible representation is $\widehat{G}$-irreducible. We do not know an example of a representation which is $\widehat{G}$-irreducible but not strongly $\widehat{G}$-irreducible.

Definition 5.8. *Let k be a finite field, and let $H \subset \widehat{G}(k)$ be a subgroup. We say that H is $\widehat{G}$-abundant if the following conditions are satisfied:*

1. *The cohomology groups $H^0(H, \widehat{\mathfrak{g}}_k)$, $H^0(H, \widehat{\mathfrak{g}}_k^\vee)$, $H^1(H, \widehat{\mathfrak{g}}_k^\vee)$ and $H^1(H, k)$ all vanish. (Here $\widehat{\mathfrak{g}}_k$ denotes the Lie algebra of $\widehat{G}_k$, and $\widehat{\mathfrak{g}}_k^\vee$ its dual.)*

2. *For each regular semisimple element $h \in H$, the torus $\mathrm{Cent}(\widehat{G}_k, h)^\circ$ is split.*

3. *For each simple $k[H]$-submodule $W \subset \widehat{\mathfrak{g}}_k^\vee$, there exists a regular semisimple element $h \in H$ such that $W^h \neq 0$ and $\mathrm{Cent}(\widehat{G}_k, h)$ is connected.*

The roles of these two definitions are as follows: the strong irreducibility of $\overline{\rho}_{\mathfrak{m}}$ allows us to cut down $\mathcal{C}^U_{G,k}$ to its finite rank $W(k)$-submodule $\mathcal{C}^U_{G,k,\mathrm{cusp}}$ using only Hecke operators (and not excursion operators). The $\widehat{G}$-abundance of $\overline{\rho}_{\mathfrak{m}}(\Gamma_{K(\zeta_\ell)})$ is used in the construction of sets of Taylor–Wiles places, which are the main input in the proof of Theorem 5.5.

If ℓ is sufficiently large, then the group $\widehat{G}(\mathbb{F}_\ell)$ is both strongly $\widehat{G}$-irreducible (inside $\widehat{G}(\overline{\mathbb{F}}_\ell)$) and $\widehat{G}$-abundant (inside $\widehat{G}(k)$, for a sufficiently large finite extension $k/\mathbb{F}_\ell$). However, it is not clear how many other families of examples there are! This motivates the following question:

Question 5.9. *Can one prove an analogue of Theorem 5.5 with weaker hypotheses? For example, can one replace conditions 3. and 4. with the single requirement that $\overline{\rho}_{\mathfrak{m}}$ is absolutely $\widehat{G}$-irreducible and ℓ is sufficiently large, relative to $\widehat{G}$?*

To weaken the '$\widehat{G}$-abundant' condition is analogous to weakening the 'bigness' condition which appeared in the first automorphy lifting theorems for unitary groups proved in Clozel, Harris, and Taylor [2008]. It seems like an interesting problem to try, in a way analogous to J. Thorne [2012], to replace this condition with the $\widehat{G}$-irreducibility of the residual representation $\overline{\rho}_{\mathfrak{m}}$.

6 Coxeter parameters

In order to apply a result like Theorem 5.5, we need to have a good supply of representations $\overline{\rho} : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{F}}_\ell)$ which we know to be residually automorphic (in the sense of arising from a maximal ideal of the excursion algebra $\mathcal{B}(U, W(k))$ acting on cuspidal automorphic forms).

Famously, Wiles used the Langlands–Tunnell theorem to prove the residual automorphy of odd surjective homomorphisms $\overline{\rho} : \Gamma_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_3)$, in order to be able to use his automorphy lifting theorems to prove the modularity of elliptic curves. Many recent applications of automorphy lifting theorems (e.g. to potential automorphy of n-dimensional Galois representations over number fields, or to the construction of lifts of residual representations with prescribed properties, as in Barnet-Lamb, Gee, Geraghty, and Taylor [2014]) have relied upon the automorphy of n-dimensional Galois representations which are induced from a character of the Galois group of a cyclic extension of numbers fields of degree n. The automorphy of such representations was proved by Arthur–Clozel, using a comparison of twisted trace formulae Arthur and Clozel [1989].

We obtain residually automorphic Galois representations from a different source, namely the geometric Langlands program. We first describe the class of representations that we use. We fix a split maximal torus $\widehat{T} \subset \widehat{G}$, and write $W = W(\widehat{G}, \widehat{T})$ for the Weyl group of $\widehat{G}$. We assume in this section that $\widehat{G}$ is simple and simply connected.

Definition 6.1. *An element $w \in W$ is called a Coxeter element if it is conjugate to an element of the form $s_1 \dots s_r$, where $R = \{\alpha_1, \dots, \alpha_r\} \subset \Phi(\widehat{G}, \widehat{T})$ is any choice of ordered root basis and $s_1, \dots, s_r \in W$ are the corresponding simple reflections.*

It is a fact that the Coxeter elements form a single conjugacy class in W, and therefore have a common order h, which is called the Coxeter number of $\widehat{G}$. They were defined and studied by Coxeter in the setting of reflection groups. Kostant applied them to the study of reductive groups Kostant [1959], and his results form the foundation of our understanding of the following definition:

Definition 6.2. *Let Γ be a group, and let Ω be an algebraically closed field. We call a homomorphism $\phi : \Gamma \to \widehat{G}(\Omega)$ a Coxeter homomorphism if it satisfies the following conditions:*

1. *There exists a maximal torus $T \subset \widehat{G}_\Omega$ such that $\phi(\Gamma) \subset N(\widehat{G}_\Omega, T)$, and the image of $\phi(\Gamma)$ in $W \cong N(\widehat{G}_\Omega, T)/T$ is generated by a Coxeter element w. We write ϕ^{ad} for the composite of ϕ with projection $\widehat{G}(\Omega) \to \widehat{G}^{ad}(\Omega)$, and T^{ad} for the image of T in $\widehat{G}^{ad}_\Omega$.*

2. *There exists a prime $t \equiv 1 \bmod h$ not dividing char Ω or $\#W$ and a primitive h^{th}-root of unity $q \in \mathbb{F}_t^\times$ such that $\phi^{ad}(\Gamma) \cap T^{ad}(\Omega)$ is cyclic of order t, and conjugation by w acts on the image by the map $v \mapsto v^q$.*[8]

We recall that if $\widehat{G} = \mathrm{SL}_n$, then $W = S_n$ and the Coxeter elements are the n-cycles. In this case the Coxeter homomorphisms appear among those homomorphisms $\Gamma \to \mathrm{SL}_n(\Omega)$ which are induced from a character of an index n subgroup. However, the above definition is valid for any simply connected simple $\widehat{G}$ and has very good properties:

Proposition 6.3. *Let $\phi : \Gamma \to \widehat{G}(\Omega)$ be a Coxeter homomorphism. Then:*

1. *ϕ is $\widehat{G}$-irreducible.*

2. *If $\phi' : \Gamma \to \widehat{G}(\Omega)$ is another homomorphism such that for all $\gamma \in \Gamma$, $\phi(\gamma)^{ss}$ and $\phi(\gamma')^{ss}$ are $\widehat{G}(\Omega)$-conjugate, then ϕ and ϕ' are themselves $\widehat{G}(\Omega)$-conjugate. In particular, ϕ is even strongly $\widehat{G}$-irreducible.*

3. *The image $\phi(\Gamma)$ is an $\widehat{G}$-abundant subgroup of $\widehat{G}(\Omega)$.*

[8]As the notation suggests, in applications we will take q to be the image in $\mathbb{F}_t$ of the cardinality of the field of scalars in K.

Now suppose that $\phi : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ is a Coxeter parameter. Then there exists a degree h cyclic extension K'/K such that $\phi(\Gamma_{K'})$ is contained in a torus of $\widehat{G}$; the homomorphism $\phi|_{\Gamma_{K'}}$ is therefore associated to Eisenstein series on $G(\mathbb{A}_{K'})$. One can ask whether it is possible to use this to obtain a cuspidal automorphic representation of $G(\mathbb{A}_K)$ (or better, a maximal ideal of the excursion algebra $\mathcal{B}_U$) to which ϕ corresponds. One case in which the answer is affirmative is as follows:

Theorem 6.4. *Let $\phi : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a Coxeter parameter such that $\phi(\Gamma_{K\cdot\overline{\mathbb{F}}_q}) \subset \widehat{T}(\overline{\mathbb{Q}}_\ell)$. Then ϕ is cuspidal automorphic, in the sense of Definition 4.5.*

Proof (sketch). Braverman–Gaitsgory construct Braverman and Gaitsgory [2002] the geometric analogue of Eisenstein series for the group G: in other words, a Hecke eigensheaf on $\mathrm{Bun}_{G,\overline{\mathbb{F}}_q}$ with 'eigenvalue' $\phi|_{\Gamma_{K\cdot\overline{\mathbb{F}}_q}}$. This Hecke eigensheaf is equipped with a Weil descent datum, which allows us to associate to it an actual spherical automorphic form $f : G(K)\backslash G(\mathbb{A}_K) \to \overline{\mathbb{Q}}_\ell$ whose Hecke eigenvalues agree with those determined by ϕ, under the Satake isomorphism. Using geometric techniques (see Gaitsgory's appendix to Böckle, Harris, Khare, and J. A. Thorne [n.d.]), one can further show that this automorphic form is in fact cuspidal. The existence of a maximal ideal in the excursion algebra $\mathcal{B}_U$ corresponding to ϕ then follows from the existence of f and the good properties of Coxeter parameters (in particular, the second part of Proposition 6.3). $\square$

To illustrate the method, here is the result we obtain on combining Theorem 6.4 with our automorphy lifting Theorem 5.5:

Theorem 6.5. *Let $\ell > \#W$ be a prime, and let $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a continuous homomorphism such that $\overline{\rho}$ is a Coxeter parameter and $\overline{\rho}(\Gamma_{K\cdot\overline{\mathbb{F}}_q})$ is contained in a conjugate of $\widehat{T}(\overline{\mathbb{F}}_\ell)$. Then ρ is cuspidal automorphic, in the sense of Definition 4.5.*

Here is a question motivated by a potential strengthening of Theorem 6.5:

Question 6.6. *Let $\phi : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ be a Coxeter parameter such that $\phi(\Gamma_{K\cdot\overline{\mathbb{F}}_q}) \subset \widehat{T}(\overline{\mathbb{Q}}_\ell)$, and let π be the everywhere unramified cuspidal automorphic representation of $G(\mathbb{A}_K)$ whose existence is asserted by Theorem 6.4. Can one show that π appears with multiplicity 1 in the space $\mathcal{Q}_{G,cusp}$?*

Taking into account the freeness assertion in Theorem 5.5, we see that a positive answer to Question 6.6 would have interesting consequences for the multiplicity of cuspidal automorphic representations.

7 Potential automorphy

We can now describe the proof of Theorem 4.6. Let us therefore choose a representation $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_\ell)$ of Zariski dense image. We must find a finite Galois extension L/K such that $\rho|_{\Gamma_L}$ is cuspidal automorphic. It is easy to reduce to the case where $\widehat{G}$ is simple and simply connected (equivalently: the group G is simple and has trivial centre), so we now assume this.

By Proposition 3.8, we can assume, after replacing ρ by a conjugate, that there is a number field E, a system $(\rho_\lambda)_\lambda$ of continuous homomorphisms $\rho_\lambda : \Gamma_{K,\varnothing} \to \widehat{G}(E_\lambda)$ of Zariski dense image, and an embedding $E_{\lambda_0} \hookrightarrow \overline{\mathbb{Q}}_\ell$ such that $\rho = \rho_{\lambda_0}$. If any one of the representations ρ_λ is automorphic, then they all are. We can therefore forget the original prime ℓ and think of the entire system $(\rho_\lambda)_\lambda$.

An application of a theorem of Larsen [1995] furnishes us with strong information about this system of representations:

Theorem 7.1. *With notation as above, we can assume (after possibly enlarging E and replacing each ρ_λ by a conjugate), that there exists a set $\mathfrak{L}$ of rational primes of Dirichlet density 0 with the following property: for each prime $\ell \notin \mathfrak{L}$ which splits in E and does not divide q, and for each place $\lambda|\ell$ of E, $\rho_\lambda(\Gamma_{K,\varnothing})$ has image equal to $\widehat{G}(\mathbb{Z}_\ell)$.*

It follows that for $\lambda|\ell$ ($\ell \notin \mathfrak{L}$ split in E), the residual representation $\overline{\rho}_\lambda$ can be taken to have image equal to $\widehat{G}(\mathbb{F}_\ell)$. It is easy to show that when ℓ is sufficiently large, such a residual representation satisfies the requirements of our Theorem 5.5.

We are now on the home straight. Using a theorem of Moret-Bailly and known cases of de Jong's conjecture Moret-Bailly [1990] and de Jong [2001], one can construct a finite Galois extension L/K and (after possibly enlarging E) an auxiliary system of representations $(R_\lambda : \Gamma_{L,\varnothing} \to \widehat{G}(E_\lambda))_\lambda$ satisfying the following properties:

1. For each prime-to-q place λ of $\overline{\mathbb{Q}}$, R_λ has Zariski dense image in $\widehat{G}(\overline{\mathbb{Q}}_\lambda)$.
2. There exists a place λ_1 such that $\overline{R}_{\lambda_1} \cong \overline{\rho}_{\lambda_1}|_{\Gamma_{L,\varnothing}}$, and both of these representations have image $\widehat{G}(\mathbb{F}_{\ell_1})$, where ℓ_1 denotes the residue characteristic of λ_1.
3. There exists a place λ_2 such that $\overline{R}_{\lambda_2}$ is a Coxeter parameter and $\ell_2 > \#W$, where ℓ_2 denotes the residue characteristic of λ_2.

The argument to prove the automorphy of $\rho|_{\Gamma_{L,\varnothing}}$ is now the familiar one. By Theorem 6.5, R_{λ_2} is cuspidal automorphic. Since this property moves in compatible systems for representations with Zariski dense image, R_{λ_1} is cuspidal automorphic. If ℓ_1 is chosen to be sufficiently large, then we can apply Theorem 5.5 to deduce that $\rho_{\lambda_1}|_{\Gamma_{L,\varnothing}}$ is cuspidal automorphic. Moving now in the compatible system containing $\rho_{\lambda_1}|_{\Gamma_{L,\varnothing}}$, we obtain finally the automorphy of the original representation $\rho|_{\Gamma_{L,\varnothing}}$, as desired.

One of the main attractions of our arguments is that they are uniform in the reductive group G. In particular, they are valid for exceptional groups. Using deformation theory, it is easy to find examples of global fields $K = \mathbb{F}_q(X)$ and continuous representations $\rho : \Gamma_{K,\varnothing} \to \widehat{G}(\overline{\mathbb{Q}}_l)$ of Zariski dense image. This gives, for example, the following simple corollary of Theorem 4.6:

Corollary 7.2. *Let G be the split simple group over $\mathbb{F}_q$ of type E_8; then the dual group $\widehat{G}$ is the split simple group over $\mathbb{Z}$ of type E_8. Let ℓ be a prime not dividing q. Then there exist infinitely pairs (K, ρ), where $K = \mathbb{F}_q(X)$ is a global field and $\rho : \Gamma_{K,\varnothing} \to E_8(\overline{\mathbb{Q}}_\ell)$ is a representation of Zariski dense image which is cuspidal automorphic, in the sense of Definition 4.5.*

Acknowledgments. I am very grateful to Chandrashekhar Khare, Michael Harris, and Vincent Lafforgue for their comments on a draft version of this article.

References

James Arthur (1989). "Unipotent automorphic representations: conjectures". *Astérisque* 171-172. Orbites unipotentes et représentations, II, pp. 13–71. MR: 1021499 (cit. on p. 442).

James Arthur and Laurent Clozel (1989). *Simple algebras, base change, and the advanced theory of the trace formula*. Vol. 120. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xiv+230. MR: 1007299 (cit. on p. 448).

Thomas Barnet-Lamb, Toby Gee, David Geraghty, and Richard Taylor (2014). "Potential automorphy and change of weight". *Ann. of Math. (2)* 179.2, pp. 501–609. MR: 3152941 (cit. on p. 448).

Gebhard Böckle, Michael Harris, Chandrashekhar Khare, and Jack A. Thorne (n.d.). "$\widehat{G}$-local systems on smooth projective curves are potentially automorphic". arXiv: 1609.03491 (cit. on pp. 434, 441, 444, 450).

A. Borel (1979). "Automorphic L-functions". In: *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*. Proc. Sympos. Pure Math., XXXIII. Amer. Math. Soc., Providence, R.I., pp. 27–61. MR: 546608 (cit. on p. 438).

A. Braverman and D. Gaitsgory (2002). "Geometric Eisenstein series". *Invent. Math.* 150.2, pp. 287–384. MR: 1933587 (cit. on p. 450).

CheeWhye Chin (2004). "Independence of l of monodromy groups". *J. Amer. Math. Soc.* 17.3, pp. 723–747. MR: 2053954 (cit. on p. 441).

Laurent Clozel, Michael Harris, and Richard Taylor (2008). "Automorphy for some l-adic lifts of automorphic mod l Galois representations". *Publ. Math. Inst. Hautes Études Sci.* 108. With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras, pp. 1–181. MR: 2470687 (cit. on p. 448).

Pierre Deligne (1980). "La conjecture de Weil. II". *Inst. Hautes Études Sci. Publ. Math.* 52, pp. 137–252. MR: 601520 (cit. on p. 440).

Fred Diamond (1997). "The Taylor-Wiles construction and multiplicity one". *Invent. Math.* 128.2, pp. 379–391. MR: 1440309 (cit. on p. 445).

Vladimir Drinfeld (n.d.). "On the pro-semisimple completion of the fundamental group of a smooth variety over a finite field". arXiv: 1509.06059 (cit. on p. 440).

A. J. de Jong (2001). "A conjecture on arithmetic fundamental groups". *Israel J. Math.* 121, pp. 61–84. MR: 1818381 (cit. on p. 451).

Bertram Kostant (1959). "The principal three-dimensional subgroup and the Betti numbers of a complex simple Lie group". *Amer. J. Math.* 81, pp. 973–1032. MR: 0114875 (cit. on p. 449).

Laurent Lafforgue (2002). "Chtoucas de Drinfeld et correspondance de Langlands". *Invent. Math.* 147.1, pp. 1–241. MR: 1875184 (cit. on pp. 434, 437, 440, 441).

Vincent Lafforgue (n.d.). "Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale". arXiv: 1209.5352 (cit. on pp. 438, 441).

M. Larsen (1995). "Maximality of Galois actions for compatible systems". *Duke Math. J.* 80.3, pp. 601–630. MR: 1370110 (cit. on p. 451).

Laurent Moret-Bailly (1990). "Extensions de corps globaux à ramification et groupe de Galois donnés". *C. R. Acad. Sci. Paris Sér. I Math.* 311.6, pp. 273–276. MR: 1071625 (cit. on p. 451).

C. Procesi (1976). "The invariant theory of $n \times n$ matrices". *Advances in Math.* 19.3, pp. 306–381. MR: 0419491 (cit. on p. 438).

R. W. Richardson (1988). "Conjugacy classes of n-tuples in Lie algebras and algebraic groups". *Duke Math. J.* 57.1, pp. 1–35. MR: 952224 (cit. on p. 439).

Jean-Pierre Serre (2005). "Complète réductibilité". *Astérisque* 299. Séminaire Bourbaki. Vol. 2003/2004, Exp. No. 932, viii, 195–217. MR: 2167207 (cit. on p. 439).

J. Tate (1979). "Number theoretic background". In: *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*. Proc. Sympos. Pure Math., XXXIII. Amer. Math. Soc., Providence, R.I., pp. 3–26. MR: 546607 (cit. on p. 436).

Richard Taylor (1991). "Galois representations associated to Siegel modular forms of low weight". *Duke Math. J.* 63.2, pp. 281–332. MR: 1115109 (cit. on p. 438).

Jack Thorne (2012). "On the automorphy of l-adic Galois representations with small residual image". *J. Inst. Math. Jussieu* 11.4. With an appendix by Robert Guralnick, Florian Herzig, Richard Taylor and Thorne, pp. 855–920. MR: 2979825 (cit. on p. 448).

Song Wang (2012). "On dimension data and local vs. global conjugacy". In: *Fifth International Congress of Chinese Mathematicians. Part 1, 2*. Vol. 2. AMS/IP Stud. Adv. Math., 51, pt. 1. Amer. Math. Soc., Providence, RI, pp. 365–382. MR: 2908081 (cit. on p. 442).

Andrew Wiles (1995). "Modular elliptic curves and Fermat's last theorem". *Ann. of Math. (2)* 141.3, pp. 443–551. MR: 1333035 (cit. on p. 445).

Received 2017-11-30.

Jack Thorne
jat58@cam.ac.uk

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (453–472)

FUNCTIONAL TRANSCENDENCE AND ARITHMETIC APPLICATIONS

Jacob Tsimerman

Abstract

We survey recent results in functional transcendence theory, and give arithmetic applications to the André-Oort conjecture and other unlikely-intersection problems.

1 Introduction

The purpose of these notes is to give an introduction to the O-minimality approach to arithmetic geometric questions in the field now referred to as "Unlikely Intersections", as well as give a brief survey of recent results in the literature. We emphasize the functional transcendence results, which are both necessary for many interesting arithmetic applications, and which we hope are of independent interest. We include some recent results on variations of hodge structures, especially as we believe these are still in their infancy and could have fantastic developments in the near future. In §2, we give a survey of the types of functional transcendence statements that have stemmed from generalizing the Ax-Schanuel theorem. In §3 we introduce o-minimality and give a sort of users manual of the main results that have been useful for these sorts of applications. We then sketch the proofs of some functional transcendence results in §4. §5 is devoted to arithmetic applications, where we explain the recent developments in the André-Oort conjecture and the Zilber-Pink conjecture. We give some sketches of proofs, but we only try to convey the main ideas, rather then give complete arguments.

2 Transcendence Theory

2.1 Classical Results. Classical transcendental number theory is largely concerned with the algebraic properties of special values of special functions. We focus first on the case of exponentiation. There are the following fundamental classical results:

MSC2010: 11G10.

Theorem 2.1. *[Lindemann-Weirstrass] Let $x_1, \dots, x_n \in \overline{\mathbb{Q}}$ be linearly independent over $\mathbb{Q}$. Then $e^{x_1}, \dots, e^{x_n}$ are algebraically independent over $\overline{\mathbb{Q}}$.*

The above result implies, for example, that $\ln q$ is transcendental for every rational number q. Baker managed to prove a similar (though weaker) result for the more difficult case of ln:

Theorem 2.2. *[Baker [1975]] Let $x_1, \dots, x_n \in \overline{\mathbb{Q}}$. If $\ln x_1, \dots, \ln x_n$ are linearly independent over $\mathbb{Q}$ then they are also linearly independent over $\overline{\mathbb{Q}}$.*

In fact Baker proved a quantitative version of the above theorem. Both of the above results are encapsulated by the following conjecture of Schanuel, which seems to encapsulate all reasonable transcendence properties of the exponential function:

Conjecture 2.1. *Let $x_1, \dots, x_n \in \mathbb{C}$ be linearly independent over $\mathbb{Q}$. Then*

$$\text{tr.deg.}_{\mathbb{Q}}\mathbb{Q}(x_1, \dots, x_n, e^{x_1}, \dots, e^{x_n}) \geq n.$$

Note that in the case where the x_i are all in $\overline{\mathbb{Q}}$ one recovers the Lindemann-Weirstrass Theorem 2.1, and in the case where e^{x_i} are all in $\overline{\mathbb{Q}}$ one recovers Baker's Theorem 2.2. Schanuel's conjecture has immediate striking implications. For instance, if one takes $n = 2$ and $\{x_1, x_2\} = \{1, \pi i\}$ then an immediate corollary is that e, π are algebraically independent over $\mathbb{Q}$.

2.2 Functional Analogue. While Schanuel's Conjecture 2.1 is still out of reach, one can get a lot more traction by considering a functional analogue. From a formal perspective, we replace the pair of fields $\mathbb{Q} \subset \mathbb{C}$ by the pair $\mathbb{C} \subset \mathbb{C}[[t_1, \dots, t_m]]$. Then one has the following theorem due to Ax, referred to as the Ax-Schanuel theorem (see Ax [1971]):

Theorem 2.3. *Let $x_1, \dots, x_n \in \mathbb{C}[[t_1, \dots, t_m]]$ have no constant term and be such that they linearly independent over $\mathbb{Q}$ Then*

$$\text{tr.deg.}_{\mathbb{C}}\mathbb{C}(x_1, \dots, x_n, e^{x_1}, \dots, e^{x_n}) \geq n + \text{rank}\left(\frac{\partial x_i}{\partial t_j}\right).$$

Example. Consider $n > m$ and let $x_i = t_i$ for $i \leq m$, and x_i to be linearly independent elements in $\mathbb{C}(t_1, \dots, t_m)$ over $\mathbb{Q}$ for $i > m$. Then it follows from Theorem 2.3 that e^{x_i} are algebraically independent over $\mathbb{C}(t_1, \dots, t_m, e^{t_1}, \dots, e^{t_m})$. It immediately follows that the set $\{e^x, x \in \mathbb{C}(t_1, \dots, t_m)\backslash\mathbb{C}\}$ is linearly independent over $\mathbb{C}(t_1, \dots, t_m)$.

We pause to explain the extra term on the right hand side in Theorem 2.3. For the moment, suppose that the x_i are convergent power series in the t_j so that the x_i can be considered as functions in the t_j. Then the x_i can be thought of as a map from $\vec{x} : \mathrm{D}^m \to \mathbb{C}^n$ for some small disk D. Now by the implicit function theorem, the dimension of $\vec{x}(\mathrm{D}^m)$ is equal to rank $\left(\frac{\partial x_i}{\partial t_j}\right)$.. Thus, if x_n contributes one to rank $\left(\frac{\partial x_i}{\partial t_j}\right)$, we can consider x_n as a formal variable over $\mathbb{C}[[x_1, \dots, x_{n-1}]]$ and then x_n, e^{x_n} are easily seen to be algebraically independent over all of $\mathbb{C}[[x_1, \dots, x_{n-1}]]$.

2.3 Geometric Formulation. Note that the above suggests a geometric reformulation of the above result. Namely, let U be the image of the map $(\vec{x}, e^{\vec{x}}) : \mathrm{D}^m \to \mathbb{C}^n \times (\mathbb{C}^\times)^n$. Note that $U \subset \Gamma$ where Γ is the graph of the exponentiation map. Then the statement of Theorem 2.3 can be reinterpreted geometrically as follows:

- $\mathrm{tr.deg.}_{\mathbb{C}}\mathbb{C}(x_1, \dots, x_n, e^{x_1}, \dots, e^{x_n})$ is the dimension of the Zariski closure of U in $\mathbb{C}^n \times (\mathbb{C}^\times)^n$
- rank $\left(\frac{\partial x_i}{\partial t_j}\right)$ is the dimension of U.
- The statement that x_i have no constant terms and are are linearly independent in $\mathbb{C}(t_1, \dots, t_m)$ over $\mathbb{Q}$, implies that the projection of U to $\mathbb{C}^n$ contains the origin and is not contained in linear subspace defined over $\mathbb{Q}$. Equivalently, the projection of U to $(\mathbb{C}^\times)^n$ is not contained in a proper algebraic subgroup.

We thus have the following geometric reformulation of Ax-Schanuel (See Tsimerman [2015] for more details):

Theorem 2.4. *Let W be an irreducible algebraic variety in $\mathbb{C}^n \times (\mathbb{C}^\times)^n$, and let U be an irreducible analytic component of $W \cap \Gamma$, where Γ is the graph of the exponentiation map. Assume that the projection of U to $(\mathbb{C}^\times)^n$ is not contained in a translate of any proper algebraic subgroup. Then*

$$\dim W = \dim U + n.$$

Even though it may seem that Theorem 2.3 is more general than the above due to the possibility of the x_i being non-convergent power series in the t_j, by formulating in terms differential fields and using the Seidenberg embedding theorem one can see that they are in fact equivalent. We list also an implication of the above theorem, as we will have use for it later. As it can be seen as analogous to the classical Lindemann-Weirstrass theorem, it has been dubbed the Ax-Lindemann-Weirstrass (or often just Ax-Lindemann) theorem by Pila:

Theorem 2.5. *Let W, V be irreducible algebraic varieties in $\mathbb{C}^n$, $(\mathbb{C}^\times)^n$ respectively such that $e^W \subset V$. Then there exists a translate S of an algebraic subgroup of $(\mathbb{C}^\times)^n$ such that $e^W \subset S \subset V$*

To deduce Theorem 2.5 one may apply the conclusion of Theorem 2.4 to the subvariety $W \times V$ of $\mathbb{C}^n \times (\mathbb{C}^\times)^n$.

2.4 Generalizations to other geometric settings. Theorem 2.4 and Theorem 2.5 are stated in the context of the exponentiation map, but it is not hard to make formal generalizations to other settings. We describe now a recipe for generalizing to other contexts. One requires the following objects:

1. Two algebraic varieties $\widehat{D}, X$, an open subset $D \subset \widehat{D}$, and a holomorphic map $\pi : D \to X$. By convention, we define an algebraic subvariety of D to be an analytic component of $D \cap V$, where V is a subvariety of $\widehat{D}$.
2. A collection S of irreducible subvarieties of X called *weakly special* varieties such that their pre-image in D contains an irreducible algebraic component.

Given the above data, we may formulate an Ax-Schanuel conjecture as follows:

Conjecture 2.2. *(Ax-Schanuel for X) Let $\Gamma \subset D \times X$ be the graph of π. Let $W \subset X$ be an irreducible algebraic variety, and U an analytic component of $W \cap \Gamma$. Then if the projection of U to X does is not contained in any proper weakly special subvariety, we have*

$$\dim U = \dim W - \dim X.$$

We proceed to give some concrete examples of this principle.

2.4.1 Abelian and Semi-Abelian Varieties. Let X be a semi-abelian variety, in other words an extension of an Abelian variety by a torus. Let the dimension of the abelian part be a and of the toric part be t, and set $g = \dim X = a + t$. Then we may take $D = \widehat{D} = \mathbb{C}^g$ and write X as D/Λ for a discrete subgroup $\lambda \subset D$ of rank $2a + t$. Now we may take the weakly special varieties to be the cosets of algebraic subgroups - which are themselves necessarily semi-abelian subvarieties. Note that this case is a direct generalization of Theorem 2.4 which we may recover by setting $a = 0$. This case was settled by Ax [1972].

2.4.2 Shimura Varieties. Let S be a Shimura variety. We do not give precise definitions in this section, referring instead to surveys such as J. Milne [n.d.] and Moonen [1998] for more details. However, such varieties are naturally quotients of Symmetric spaces D by arithmetic groups, and moreover the spaces D can be identified as a quotient of real lie groups $D \cong G(\mathbb{R})/K$ for G a semisimple lie group defined over $\mathbb{Q}$, and K a maximal compact subgroup of G. This means that $G(\mathbb{R})$ acts on D, and the weakly special varieties can be characterized as the images of orbits $H(\mathbb{R}) \cdot v$, where $H \subset G$ is a semisimple lie subgroup defined over $\mathbb{Q}$ such that the orbit $H(\mathbb{R}) \cdot v$ is complex analytic. Thus, even though D and S do not form groups, the machinery of group theory is still very much present in this setting, though of course the groups are non abelian making this setting significantly more difficult then the abelian and semi-abelian case.

Example. For a positive integer n, one may take $D = \mathbb{H}^n, S = Y(1)^n$ and $\pi : D \to S$ be the j map, where $Y(1)$ is the (coarse) moduli space of complex elliptic curves and $\mathbb{H}$ is the usual upper-half plane. The weakly special shimura varieties V can be described very simply as being imposed by one of 2 types of conditions:

- One may impose a co-ordinate of S to be a constant
- One may insist that, for a fixed positive integer N, tow co-ordinates x_i, x_j of S correspond to elliptic curves which are related by a cyclic isogeny of degree N.

The above 2 operations may yield varieties that are not irreducible, so one should also be allowed to take irreducible components. This yields for a very nice combinatorial description of weakly pecial subvarieties which becomes significantly more complicated for other Shimura varieties, but in practise the explicitness of the description is rarely essential to proofs. In this context, the Theorem 2.5 was proven by Pila [2011] in his groundbreaking unconditional proof of André-Oort for $X(1)^n$, and the Theorem 2.4 was proven by Pila and Tsimerman [2016].

The general case of Theorem 2.5 was proven in Klingler, Ullmo, and Yafaev [2016] after partial progress was done by Ullmo and Yafaev [2014] for compact S, and by Pila and Tsimerman [2014] for $\mathfrak{A}_g$. The general case of Theorem 2.4 was recently announced by Mok, Pila, and Tsimerman [2017]. One may also generalize to *Mixed Shimura Varieties*, where Theorem 2.5 was proven by Gao [2017].

2.4.3 Hodge Structures. There is another generalization one may make, and that is to the setting of Hodge Structures. This setting is slightly more complicated than what we have defined so far and it doesn't exactly fit into our setup, for reasons we will describe. Nevertheless, it is important for arithmetic reasons that we will mention in later sections. For general background on Hodge Structures, we refer the reader to Voisin [2002]. We give a quick definition here.

Definition. *An integral hodge structure of weight m and dimension n consists of:*

- *A free abelian group L of dimension n*
- *An integer-valued non-degenerate quadratic form Q on L satisfying $Q(v,w) = (-1)^n Q(w,v)$*

- *A* ***Hodge decomposition*** *of complex vector spaces* $L \otimes \mathbb{C} = \oplus_{p+q=n} H^{p,q}$ *such that* $\overline{H^{p,q}} = H^{q,p}$ *and* $i^{p-q} Q(v, \bar{w})$ *is positive definite on* $H^{p,q}$.

The numbers $h^{p,q} := \dim H^{p,q}$ *are called the* Hodge numbers.

If one fixes a weight, dimension, hodge numbers, and isomorphism class of (L, Q) one obtains a complex analytic space X which parametrizes hodge structures up to isomorphism. Moreover, if one further fixes a basis for L, one obtains an open subset D of a complex Grassmanian manifold, and a holomorphic covering map $D \to X$ with Monodromy $G(\mathbb{Z})$, where $G = \mathrm{Aut}(L, Q)$. Moreover, $G(\mathbb{R})$ acts transitively on $D(\mathbb{R})$ with compact stabilizer, so we obtain a picture quite similar to the one which occurs for Shimura varieties. Indeed, by including some extra data one may recover all Shimura varieties as moduli of Hodge structures.

The reason that this setting presents significant additional complication is that for 'most' choices of Hodge numbers, Carlson and Toledo [2014] showed that X cannot be endowed with the structure of an algebraic variety. In our specific context, this makes formulating a transcendence conjecture difficult!

To resolve this problem, we note that a primary motivating reason for studying hodge structures is that for smooth projective varieties Y and a positive integer m, the cohomology group $H^m(Y, \mathbb{C})$ can naturally be given a hodge structure, with the integer lattice coming from Betti cohomology, the Hodge decomposition coming from Dolbeaut Cohomology, and the quadratic form coming from the cup product. This means that even though the moduli space X is not algebraic, for any family of smooth algebraic varieties over a base B such that the fibers have the right Hodge numbers, we get a *period map* $B \to X$. These period maps give us an algebraic structure to work with (namely, that on B) and so allows us to formulate a version of Theorem 2.4. Such a statement was conjectured by Klingler[1] in Klingler [n.d.], and was proven by Bakker and Tsimerman [2017]. The proof follows closely the structure of Mok, Pila, and Tsimerman [2017], with the primary difference being a new "volume-growth" inequality for moduli of Hodge structures in Griffiths-Transverse directions.

3 O-minimality

3.1 Definitions and Introduction. In the course of proving Theorem 2.4 as well as all its generalizations, one naturally deals with functions that are not algebraic. However, the set of all complex analytic functions can be too unwieldy, so it is natural to look for an intermediate category of functions in which to work. It turns out that one such theory which works particularly well for this class of problems is that of o-minimal structures.

[1] In fact, Klingler conjectures much more specific results

This allows us to work with enough functions to be able to talk about the transcendental covering maps in Theorem 2.4, while still maintaining many of the nice properties that algebraic functions possess.

For our purposes, a **Structure** $\mathcal{S}$ is a collection of sets $S_n \subset 2^{\mathbb{R}^n}$, where the elements of S_n are subsets of $\mathbb{R}^n$, such that the following properties hold:

- S_n is a boolean algebra
- $S_n \times S_m \subset S_{m+n}$
- If we let $\pi : \mathbb{R}^n \to \mathbb{R}^j$ be co-ordinate subspace projection, and $A \in S_n$, then $\pi(A) \in S_j$.
- The set $\{(x, x), x \in \mathbb{R}\}$ is in S_2, and the sets $\{(x, y, x + y), x, y \in \mathbb{R}\}, \{(x, y, xy), x, y, \in \mathbb{R}\}$ are in S_3.

We further say that $\mathcal{S}$ is *o-minimal* if S_1 consists precisely of finite unions of open intervals and points. We say that a set $Z \subset \mathbb{R}^n$ is *definable* in § if $Z \in S_n$, and we say that a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is definable in $\mathcal{S}$ if its graph is.t turns out that o-minimal structures have a myriad of useful properties. For example, any set definable in an o-minimal structure has finitely many connected components, has a well-defined dimension, and is almost everywhere differentiable. For an introduction to the theory, see van den Dries [1998].

3.2 Examples of o-minimal structures. It follows from the definitions that the smallest possible structure is the structure $\mathbb{R}_{sa}$ which contains all semi-algebraic sets. It follows from the Tarski-Seidenberg theorem $\mathbb{R}_{sa}$ is o-minimal. It is highly non-trivial to prove that any enlargements are o-minimal.

- Gabrièlov [1968] prove that the structure $\mathbb{R}_{an}$, which is defined as the smallest structure containing all subanalytic functions is definable. Recall that a subanalytic function is a function $f : T \to \mathbb{R}$ where $T \subset \mathbb{R}^n$ is a compact ball such that f extends to an analytic function on an open neighbourhood of T.
- Building on work of Khovanskiĭ [1991], Wilkie [1996] proved that the structure $\mathbb{R}_{exp}$, which is defined as the smallest structure containing the graph of the **real** exponential function[2], is o-minimal.
- The strucure $\mathbb{R}_{an,exp}$ is defined to be the smallest structure containing $\mathbb{R}_{an}$ and $\mathbb{R}_{exp}$, and the o-minimality of this structure was shown by van den Dries and Miller

[2] The complex exponential function has countable pre-images, and so cannot be o-minimal.

[1994b]. Note that the structure generated by two o-minimal structures need not be o-minimal Rolin, Speissegger, and Wilkie [2003], so this is by no means a trivial theorem. The structure $\mathbb{R}_{an,exp}$ turns out to be large enough to encompass most functions that are needed for arithmetic applications, and so this is the structure we will ultimately work in.

3.3 Counting Rational Points in Definable Sets. One very common and useful heuristic in number theory is that sets contain "few" integer/rational points unless they have some kind of "reason" for doing so. As such, one would expect that Transcendental subvarieties contain few such points. However, one must be careful to avoid dealing with sets that are too unwieldy. For example, the graph of the function $\sin(\pi x)$ contains every integer point of the x-axis, and is quite transcendental. It turns out that for o-minimal sets such a thing can't happen, as was shown by Bombieri and Pila [1989]. To state their theorem, let us define the *height* of a rational number $x = a/b$ with $\gcd(a, b) = 1$ to be $H(x) = \max(|a|, |b|)$ and the height of a rational point $x = (x_1, \ldots, x_n)$ to be $H(x) = \max_i(H(x_i))$. For a subset $\mathbb{Z} \subset \mathbb{R}^n$ we define the counting function of Z by

$$N(Z, T) := \#\{x \in Z \cap \mathbb{Q}^n \mid H(x) \leq T\}.$$

Theorem 3.1. *Bombieri and Pila [ibid.] Let $Z \subset \mathbb{R}^2$ be a compact real-analytic transcendental curve. Then*

$$N(Z, T) = T^{o(1)}.$$

In other words, the number of points on Z grows subpolynomially.

The proof of the above theorem uses the determinant method, whereby one uses the rational points to form a determinant that has a lower bound stemming from arithmetic, and an upper bound stemming from geometry. One would like to generalize the above theorem to higher dimensions, but some care is required stemming from the fact that a transcendental surface could easily contain an algebraic curve, or even a line, and thus contain lots of rational points. As such, we define Z^{alg} to be the union of all semi-algebraic curves contained in Z. Then Pila and Wilkie [2006] prove the following higher dimensional generalization:

Theorem 3.2. *Pila and Wilkie [ibid.]*

Let $Z \subset \mathbb{R}^n$ be definable in an o-minimal structure. Then

$$N(Z - Z^{alg}, T) = T^{o(1)}.$$

In other words, the number of points on $Z - Z^{alg}$ grows subpolynomially.

One may obtain the same quality bound for counting not only rational points, but algebraic points of a bounded degree over $\mathbb{Q}$. The proof of the above theorem proceeds roughly as follows: One applies the determinant method to show that the rational points in Z lie in the intersections $Z \cap V$ of Z with "few" hypersurfaces V of small degree. One then wishes to apply the theorem inductively on dimension. The key to doing this is a parametrization theorem which means that the intersections $Z \cap V$ can be parametrized by finitely many maps with uniformly bounded derivatives as V varies in the family of all hypersurfaces of a given degree. This is where the o-minimality is crucial to the argument.

3.4 Tame complex geometry. We first say a word about extending the notion of definability to sets that aren't subsets of $\mathbb{R}^n$. First, if Z is a subset of $\mathbb{C}^n$ one may use the identification $\mathbb{C}^n \cong \mathbb{R}^{2n}$ to talk about the definability of Z. Moreover, we may talk about definable manifolds, by insisting on *finite* open covers that have an isomorphism onto a definable subset of $\mathbb{R}^m$ such that the transition maps are definable. In particular, this gives every complex algebraic variety the structure of a definable (in any structure) manifold by taking a finite affine open cover,.

Peterzil and Starchenko have a series of works Peterzil and Starchenko [2010] where they develop complex geometry in an o-minimal setting[3] where they prove tameness of complex analytic sets that are definable in an o-minimal setting. One particularly robust and applicable result to the arithmetic setting is the following version of Chows theorem, which works for any algebraic variety, not just proper varieties!

Theorem 3.3. *Peterzil and Starchenko [ibid.] Let V be a complex algebraic variety, and $S \subset V$ be a closed, complex-analytic subset, which is definable in an o-minimal structure. Then S is an algebraic subvariety of V.*

Note that complex analytic, closed subvarieties of proper algebraic varieties are definable in $\mathbb{R}_{sa}$, so Theorem 3.3 has the usual Chow theorem as an immediate consequence. Note that the above theorem is not stated in the above form in Peterzil and Starchenko [ibid.], but is easily deducible from those results. The deduction is spelled out in a few places, for example in Mok, Pila, and Tsimerman [2017].

3.5 Fundamental Domains and Definability of Uniformization Maps. In §1, we introduced the setting where we have transcendental Uniformization maps $\pi : D \to S$, where S, D are open subsets of complex algebraic varieties. One may hope for the maps π to be definable in $\mathbb{R}_{an,exp}$ with respect to the natural definable structure on D, S. However, the inverse images of points under these maps π are countable, discrete sets, and

[3] In fact, they in the more general setting of a totally-ordered field instead of $\mathbb{R}$

therefore cannot be definable in *any* o-minimal structure! Instead, what one does is restrict to a fundamental domain $\mathfrak{F} \subset D$. That is, one looks for a definable subset $\mathfrak{F}$ such that $\pi \mid \mathfrak{F}$ is an isomorphism onto S, and then asks whether that isomorphism is definable. Note that this is dependant on the fundamentalddomain that one chooses. Below we give some relevant examples:

- In the case of the exponential function $\pi : \mathbb{C}^n \to \mathbb{C}^{\times n}$, we use the fundamental domain

$$\mathfrak{F} = \{(z_1, \ldots, z_n), \Im(z_i) \in [0, 2\pi]\}.$$

  In the real coordinates $z = x + iy$ the function e^z becomes $e^x \cos(y) + ie^x \sin(y)$. Now since we are restricting y to be in a bounded interval, $\cos y, \sin y$ restricted to this interval are definable in $\mathbb{R}_{an}$ and thus $\pi \mid \mathfrak{F}$ is definable in $\mathbb{R}_{an,exp}$.

- In the case of an abelian variety, or indeed any case where S is compact, one may take for $\mathfrak{F}$ any bounded fundamental domain, and $\pi \mid \mathfrak{F}$ will be definable in $\mathbb{R}_{an}$.

- In the case of the j-function, $j : \mathbb{H} \to \mathbb{C}$, we may use the usual fundamental domain given by $\mathfrak{F} = \{z \in \mathbb{H}, \Re e(z) \in [-1/2, 1/2), |z| \geq 1\}$, and use the Laureny expansion of j in terms of $e^{2\pi iz}$ to see that $j \mid \mathfrak{F}$ is definable in $\mathbb{R}_{an,exp}$.

- In the more general case of a Shimura variety $\pi : D \to S$, where $D \cong G(\mathbb{R})/K$ and $S = D/G(\mathbb{Z})$, one may use the Iwasawa decomposition $G = NAK$ to make a Siegel set for $G(\mathbb{R})$, and translate that to a fundamental domain $\mathfrak{F}$. It is substantially more difficult to show that $\pi \mid \mathfrak{F}$ is definable in $\mathbb{R}_{an,exp}$. It was done for the moduli space $\mathfrak{A}_g$ of principally polarized abelian varieties by Peterzil and Starchenko [2013] and in general by Klingler, Ullmo, and Yafaev [2016]. This result was generalized to the Mixed Shimura case by Gao [2017].

- As mentioned before, the case $\pi : D \to X$ where D is a period domain and X parametrizes hodge structures of given hodge numbers is more difficult. Since X does not typically admit any algebraic structure it does come equipped with a definable structure either. Nevertheless, one may use any definable fundamental domain $\mathfrak{F}$ endow X with a definable $\mathbb{R}_{an,exp}$-structure. However, if we proceed with this setup, one wants the following natural property to be satisfied: Given a variation of hodge structures over an algebraic base B, one would like the period map $\psi : B \to X$ to be definable with respect to this structure. In forthcoming work of the author with Bakker, it is shown that if one uses a definable fundamental domain $\mathfrak{F}$ coming from a Siegel set, then the period maps are indeed definable.

4 Proofs of Functional Transcendence results

We attempt in this section to give a brief idea of how these results are proven. We first describe the proof of Theorem 2.5.

4.1 Ax-Lindemann Theorems. We restrict ourselves for expository purposes to the setting of a torus, and briefly explain how to adapt the methods to the setting of a Shimura variety. So suppose that $\pi : \mathbb{C}^n \to \mathbb{C}^{\times n}$ is the exponential map, and let $V \subset \mathbb{C}^n$, $W \subset \mathbb{C}^{\times n}$ be algebraic varieties such that $\pi(V) \subset W$. We pick our usual definable fundamental domain $\mathfrak{F}$, and we let $W' = (\pi \mid \mathfrak{F})^{-1}(W)$. Crucially for us W' is definable. The key idea of the proof lies in considering the following set:

$$I = \{t \in \mathbb{C}^n \mid (V + t) \cap \mathfrak{F} \subset W'.$$

It is easy to see that I is definable in $\mathbb{R}_{an,exp}$. Moreover, since $\pi^{-1}(W)$ is invariant under the monodromy group $\mathbb{Z}^n$, it follows that I contains all those elements $t \in \mathbb{Z}^n$ such that the V intersects $\mathfrak{F} - t$. There must be polynomially many of these elements (in fact, at least linearly many) and thus, by the Counting Theorem 3.2 we can conclude that I contains semialgebraic curves. It follows that there is a complex algebraic curve C such that $V + c \subset W'$ for all $c \in C$.

Now we may try to replace V by $V \cap V + C$ and use an induction argument on $\dim V - \dim W$. This will work unless V is invariant under C. If this is the case for all curves C, it would imply that the intersection of V with $\mathfrak{F} + t$ all 'look the same', or in other words that $\pi(V) = \pi(V \cap \mathfrak{F})$. But this implies that $\pi(V)$ is definable in $\mathbb{R}_{an,exp}$, and thus by the definable chows Theorem 3.3, we see that $\pi(V)$ is algebraic. The proof now follows from momodromy arguments.

To generalize the above to the context of Shimura varieties, one encounters two difficulties:

- One needs an additional argument to show that there are many elements $g \in G(\mathbb{Z})$ sich that V intersects $g \cdot \mathfrak{F}$. In the general case, one now uses a hyperbolic volume argument due to Hwang and To [2002], that says that in hyperbolic balls of radius r, the volume of a complex analytic ball must grow at least exponentially in r. By contrast, one can show using definability and siegel set arguments that the volume in any fundamental domain is bounded by a constant. Thus, any curve must pass through at least c^r fundamental domains of distance r away, and then one relates the distance of a fundamental domain $g \cdot \mathfrak{F}$ to the height of g.

- Additional care must be taken due to the non-abelian nature of the groups involved. This is indeed a difficulty, but by setting things up correctly the argument goes through.

4.1.1 Ax-Schanuel. One may use a similar setup to the above, but now we only know that $\dim(V \cap W') > \dim V + \dim W' - n$. Nontheless, we may still define

$$I = \{t \in \mathbb{C}^n \mid \dim\left((V+t) \cap \mathfrak{F} \cap W'\right) = \dim(V \cap W')$$

and conclude that I contains many points. In fact, in this setting the argument can pushed through (see Tsimerman [2015]). However, in the Shimura setting, due to the extremely non-abelian nature of the groups involved and the fact that $V \cap \pi^{-1}(W)$ is not algebraic either in D, not once its pushedforward to X, the argument seems difficult to push through. However, there is a brilliant idea of Mok [n.d.] [4] which provides a great help in this context.

Mok realizes that by working in W, one can formulate the condition of the existence of V algebraically through a differential equation, even though the map π is extremely transcendental! This means that for a given W, if one V exists giving an excessively large dimension, then there must be a whole family of such V. With this extra freedom in hand to vary V, the argument goes through.

5 Arithmetic applications

We will discuss some problems that typically fall under the ”atypical intersection” umbrella.

5.1 Langs conjecture. Consider again the setting of the torus $X = \mathbb{C}^{\times n}$. The torsion points - points whose co-ordinates are all roots of unity - are distinguished algebraic points in X, and it is natural to ask which algebraic subvarieties contain infinitely many torsion points. In fact, it turns out to be a better (and essentially equivalent question) to ask which irreducible subvarieties contain a Zariski-dense set of torsion points. It is clear that subtori do, and in fact so do coset of a subtorus by a torsion point. The converse was conjectured by Lang [1966] and proven by Raynaud [1983a,b]:

Theorem 5.1. *Let $V \subset \mathbb{C}^{\times n}$ be an irreducible subvariety containing a Zariski dense set of torsion points. Then V is a coset of a subtorus of $\mathbb{C}^{\times n}$ by a torsion point.*

Recall that we defined the weakly special subvarieties to be cosets of subtori of X. The reason we used that strange terminology, is that we say that a *special subvariety* of X is translate of a torus by a torsion point. The above theorem is easily seen to be equivalent to the following statement: *The Zariski closure of an arbitrary union of special subvarieties is a finite union of special subvarieties*.

[4] In fact, Mok uses this idea in Mok [n.d.] to prove the Ax-Lindemann conjecture for all rank-1 quotients of hyperbolic space, even non-arithmetic ones!

The statement of Lang's conjecture can be generalized to the setting of abelian varieties almost verbatim, where we replace the word ”subtorus” by ”algebraic subgroup”. The resulting conjecture is known as the Manin-Mumford conjecture, and was first proven by Raynaud [1983a,b]. Below we will give a proof of Lang's conjecture using the ideas we have developed in the previous section with o-minimality and functional transcendence, following Pila and Zannier [2008].

5.2 André-Oort conjecture. Let S be a shimura variety, and $\pi : D \to S$ be its covering by the corresponding symmetric space. There is a natural $\overline{\mathbb{Q}}$ structure on the variety S, and there are distingiushed $\overline{\mathbb{Q}}$ points on S called *CM points*. We call $V \subset S$ a *special subvariety* if V is a weakly special subvariety which contains at least one CM point, which will inn fact force V to contain a Zariski-dense set of CM points. If $S = Y(1)$ is the moduli space of elliptic curves, then the CM points correspond to those complex elliptic curves E with extra endomorphisms, so that $\mathbb{Z} \subsetneq \mathrm{End}(E)$. More generally, if $S = \mathfrak{A}_g$ is the moduli space of principally polarized abelian varieties of dimension g, than the CM points correspond to those Abelian varieties A such that the endomorphism algebras $\mathrm{End}(A) \otimes \mathbb{Q}$ contain a field of degree $2g$ over $\mathbb{Q}$. This can be intuitively thought of as saying that A has “as many symmetries as possible”. Note that it is not immediately obvious that such abelian varieties are even defined over $\overline{\mathbb{Q}}$!

The André-Oort conjecture is the natural generalization of Lang's conjecture to this setting:

Conjecture 5.1. *Let $V \subset S$ be an irreducible subvariety containing a Zariski dense set of torsion points. Then V is a special subvariety.*

There has been much work on the AO conjecture. It was first proven unconditionally for $Y(1)^2$ by André, and later proven in generality but conditionally on the generalized Riemann hypothesis in Ullmo and Yafaev [2014], building on an idea of Edixhoven [2005] who handle the case of $Y(1)^2$ conditionally. Later, Pila adapted his method with Zannier to prove the AO conjecture unconditionally for $Y(1)^n$, and it was recently proven for $\mathfrak{A}_g$. The general case remains open

5.3 Crucial Ingredient: Galois Orbits.

5.3.1 Tori. A crucial ingredient in the Pila-Zannier approach to the special point problems that we have discussed is the ability to prove lower bounds for Galois orbits of special points. In the setting of Lang's conjecture, the special points are simply torsion points of the torus, so in this setting the relevant galois action is the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the roots of unity. By Class field theory for $\mathbb{Q}$, we know that this action is transitive on points of exact order n for each n, so we know precisely how large the Galois orbits are.

5.3.2 Abelian Varieties. In the setting of an Abelian variety A over $\mathbb{Q}$, it is much more difficult to get a handle on the action on the torsion points of A. Given a point P of order n, the lower bound that one needs is $[\mathbb{Q}(P) : \mathbb{Q}] \gg n^\delta$ for some positive integer $\delta > 0$. There are a few was to proceed here. The Galois action on torsion points of A gives rise to a morphism $\rho_A : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Gsp}_{2d}(\widehat{\mathbb{Z}})$ where $d = \dim A$, and the action on torsion points can be read off easily from this action. It is conjectured that the image is open in the $\widehat{\mathbb{Z}}$ points of the Mumford-Tate group. While this is still open, it follows from work of Serre that that the image contains a power of the center, $(\widehat{\mathbb{Z}}^\times)^m$, for some positive integer m. This immediately implies that the orbit of any torsion point is of size at least $|(\mathbb{Z}/n\mathbb{Z})^{\times m}| \gg n^{1-o(1)}$, which is sufficient.

There is also an analytic approach by D. W. Masser [1977] which yields this result in a form more suitable for studying families of Abelian varieties.

5.3.3 Shimura Varieties. In the setting where X is a Shimura variety and $p \in X$ is a CM point, one may use the theory of complex multiplication developed by Shimura, Taniyama, and others to relate the size of the Galois orbit of p to class groups of number fields. For example, if $X = X(1)$ and p corresponds to an elliptic curve E_p with endomorphism ring the ring of integers in $K = \mathbb{Q}(\sqrt{D})$ then the size of the Galois orbit of p is equal to the class number of K, which is asymptotic to $|D|^{\frac{1}{2}+o(1)}$. In general, there are two naturally associated tori S, T over $\mathbb{Q}$ such that the size of the Galois orbit of p is the image of the class group of S in the class group of T. Class groups of Tori can be defined naturally just as for number fields (see Shyr [1977]) and the sizes of the class groups satisfy an asymptotic Brauer-Siegel formula which gives us very precise control. However, the challenge comes from the fact that these isogenies can kill torsion of low order, and it is very difficult to obtain unconditional upper bounds on low-order torsion in class groups of number fields. In particular, it is a conjecture that for a number field K of discriminant D, fixed degree n over $\mathbb{Q}$, and a positive integer m that $|CL(K)[m]| = |D|^{o(1)}$, and yet one cannot in most cases even beat the trivial bound $|D|^{\frac{1}{2}+o(1)}$ given by Brauer-Siegel! For results in this direction see Bhargava, Shankar, Taniguchi, Thorne, Tsimerman, and Zhao [2017], Ellenberg and Venkatesh [2007]. However, we cannot even show that the class group of imaginary quadratic fields are not all mostly 5-torsion! If one assumes GRHOne can show something in this direction by using GRH to produce small split primes, and this is the primary reason that André-Oort is only known unconditionally under GRH. Nevertheless, one may push these methods to prove AO unconditionally for $\mathcal{A}_g$ for $g \le 6$. See Tsimerman [2012],Ullmo and Yafaev [2015].

5.3.4 The case of $\mathcal{A}_g$. The required lower bounds were recently established for $X = \mathcal{A}_g$ using methods different from the above in Tsimerman [2018]. As a corollary, one derives the following result (which seems to be of the same level of difficulty), which was not preivously known:

Theorem 5.2. *For each positive integer g, there are finitely many CM points in $\mathcal{A}_g(\mathbb{Q})$.*

We briefly describe the proof. Let x be a CM point in $\mathcal{A}_g$. Then x occurs in a finite collection C of those CM points with the same endomorphism ring and CM type as x. In the case of elliptic curves $g = 1$ the set C is a single Galois orbit, but for larger g that is not usually the case. Moreover, the set C is acted on by the Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and the orbits all have the same size. So if x is defined over $\mathbb{Q}$, then all the points in C are defined over $\mathbb{Q}$. Moreover, all the points in C are isogenous. Now one uses a famous theorem of D. Masser and Wüstholz [1993]:

Theorem 5.3. *Let A, B be isogenous abelian varieties over some number field K. Then the degree of the smallest isogeny N between them satisfies $N \leq \max(h(A), [K : \mathbb{Q}])^{c_g}$ where $h(A)$ denotes the Faltings height of A, and $c_g > 0$ is a positive constant depending only on g.*

In other words, if two abelian varieties are isogenous, then there must exist an isogeny between them whose degree is not too large. Applying the above theorem to any two points in C and using our assumption that all the points in C are defined over $\mathbb{Q}$, it follows that they all have isogenies between them of degree at most $h(A)^{c_g}$. However, there are only polynomially many isogenies of degree N that one can take, so if $h(A)$ is sufficiently small one obtains a contradiction.

Now, in general heights of abelian varieties can be quite hard to get a handle on. However, for CM abelian varieties A, Colmez [1993] has a beautiful conjecture computing the Faltings height of A in terms of certain L-values at 1 of Artin representations. This conjecture combined with standard estimates on L functions implies the desired upper bound on $h(A)$. While the Colmez conjecture is still open in general, it was recently proven independently in Andreatta, Goren, Howard, and Madapusi Pera [2018], Yuan and Zhang [2018] that if one averages over a finite family of CM types, the Colmez conjecture is true. This finite average has minimal effect from an analytic standpoint, so is enough to complete the proof.

5.4 Proof of the André-oort conjecture. Once one has the required Galois orbit lower bounds and Functional Transcendence results at ones disposal, the proofs of the André-Oort conjecture and the Lang conjecture proceed among essentially identical lines, so we give them both at once using the language of special varieties. So suppose that $\pi : D \to X$

is our covering map, $V \subset X$ is an algebraic variety and V contains a Zariski-dense set of special points. It follows that V is defined over $\overline{\mathbb{Q}}$, and thus over a number field. For simplicity of exposition we assume that V is defined over $\mathbb{Q}$, though this minimally affects the proof. Let x_i be a sequence of CM points which is Zariski-dense in V. Let $\mathfrak{F} \subset D$ be a standard fundamental domain, and consider the pre-images y_i of the x_i under $\pi \mid \mathfrak{F}$. It turns out that the y_i are all defined over number fields of bounded degree over $\mathbb{Q}$. For example, in the case of Lang's conjecture, the pre-image under $z \to e^{2\pi i z}$ of the torsion points the rational numbers , and if one restricts to a suitable fundamental domain one obtains the rational numbers between 0 and 1. Moreover, the Galois lower bounds imply that the heights of these numbers satsify $H(y_i) \ll |\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})x_i|^{\delta}$ for some fixed positive constant δ. Since V contains all of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})x_i$, it follows that the pe-image W of V under π contains a lot of rational points. By the Counting Theorem 3.2 it follows that W must contain algebraic subvarieties containing all but finitely many of these special points. By Theorem 2.5 and its generalizations, it follows that these algebraic subvarieties must be pre-images of special varieties contained in V. We've thus succeeded in showing that all but finitely many special points in V are contained in higher dimensional special subvarieties. At this point, an induction argument using to finish the proof. We don't give the argument since it requires some deeper analysis using definability in o-minimal structures, and instead refer the interested reader to Pila and Zannier [2008] and Tsimerman [2015].

5.5 The Zilber-Pink Conjecture. Let us return to the setting of a Torus, and consider a proper subvariety $V \subset \mathbb{C}^{\times n}$. For any special subvariety (coset of a subtorus by a torsion point) T, naive dimension theory suggests that the dimension of $V \cap T$ is $\dim V + \dim T - n$. Thus, whenever $\dim(V \cap T) + \dim n > \dim V + \dim T$ we call $V \cap T$ an *unlikely intersection* for V. Notice that if T is a point, then V only intersects T if V contains T, in which case the intersection will be unlikely. Thus this concept generalizes the special point problems studied above. Of course, unlikely intersections can be easily constructed. For example, one may take V to be a subvariety of a subtorus. Then all of V is an unlikely intersection for V! More subtly, one may take any codimension ≥ 2 special variety T, take a codimension 1 subvariety $U \subset T$ and then arbitrarily take V to be another variety containing U as a divisor. However, one has the following conjecture, made by Bombieri-Masser-Zannier:

Conjecture 5.2. *Bombieri, D. Masser, and Zannier [1999] and Zilber [2002]*

Let $V \subset \mathbb{C}^{\times n}$ be a proper subvariety. Then there are finitely many unlikely intersections for V which are maximal *under inclusion.*

One may of course easily generalize to the setting of abelian varieties or (mixed) Shimura varieties, and it is in this general setting that the Zilber-Pink conjecture occurs. One may generalize the Pila-Zannier method to this setting, but this conjecture is substantially more difficult than the corresponding one for special point problems. For one thing, the required functional transcendence input is the Ax-Schanuel Theorem rather then the easier Ax-Lindemann theorem. However, this has been established in (essentially) complete generality so is no longer an obstruction. However, the lower bounds for Galois orbits that are required seem completely out of reach in general. In the André-Oort conjecture, we are interested in lower bounds of CM points, for which we have all the understanding provided by the theory of complex multiplication, and even here the problem is not solved. For the Zilber-Pink conjecture, one must understand Galois orbits of $V \cap T$, and these have no discernible special structure.

Nonetheless, there are impressive partial results. In the setting of a Torus, the result is known if V is a curve by Bombieri, D. Masser, and Zannier [1999]. In the shimura variety setting, It is proven by Habegger-Pila that if $C \subset Y(1)^n$ is a curve where the degrees of the n projections $C \to Y(1)$ are all different, then the Zilber-Pink conjecture holds. Their proof uses the Masswer-Wüstholz theorem in a very clever way to get prove the required Galois lower bounds. This result was recently partially generalized to certain curves in $\mathfrak{A}_g$ by Orr [2017].

5.6 Results on integral points. In recent, as of yet unpublished work, Lawrence-Venkatesh have come up with a new method by which to use transcendence results to prove powerful finiteness results concerning *integer points* on Varieties. There are sometimes called Shafarevich-type theorems after Shafarevich's theorem that there do not exist elliptic curves over $\operatorname{Spec} \mathbb{Z}$. Briefly, their idea is as follows. Consider a smooth, projective family $Y \to B$ over some algebraic variety $B/\mathbb{Q}$, such that some fibral cohomology group $H^n(Y_b, \mathbb{C})$ is non-zero and the corresponding period map $\psi : B \to X = D/\Gamma$ is not constant. Then one may use global results on Faltings to show that the Galois representations $\rho_b : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to H^n(Y_b, \mathbb{Q}_p)$ occur in finitely many isomorphism classes, as b varies over the integer points $B(\mathbb{Z})$. The reason one requires integer rather than rational points is so that one may control how many primes of bad reduction ρ_b has. Now consider a "p-adic lift" of ψ, which looks like $\tilde{\psi} : B(\mathbb{Q}_p) \to D(\mathbb{Q}_p)$.By results of p-adic hodge theory, the finiteness of Galois representations of the ρ_b implies that $\tilde{\psi}(B(\mathbb{Z}))$ is contained in an algebraic subvariety H of D. Now if $B(\mathbb{Z})$ is infinite, or Zariski-dense, one obtains a p-adic violation of a version of the Ax-Schanuel Theorem 2.4. In fact, this last part is not a complication, since one may formally deduce the p-adic Ax-Schanuel theorem from the complex version proven in Bakker and Tsimerman [2017]

Of course, we are skirting a myriad of complexities, but they can already prove the Mordell conjecture[5] using their methods - for which they do not require the Ax-Schanuel theorem, so it seems quite possible that this method has the potential to prove much more.

[5] The paper in its current form only handles certain cases, but they now claim the full result

Acknowledgments. It is a pleasure to thank Jonathan Pila and Peter Sarnak for many discussions regarding the above topics, and for teaching me much of what I know about the subject.

References

Yves André (1992). "Mumford-Tate groups of mixed Hodge structures and the theorem of the fixed part". *Compositio Math.* 82.1, pp. 1–24. MR: 1154159.

Fabrizio Andreatta, Eyal Goren, Benjamin Howard, and Keerthi Madapusi Pera (2018). "Faltings heights of abelian varieties with complex multiplication". *Ann. of Math. (2)* 187.2, pp. 391–531. arXiv: 1508.00178. MR: 3744856 (cit. on p. 467).

James Ax (1971). "On Schanuel's conjectures". *Ann. of Math. (2)* 93, pp. 252–268. MR: 0277482 (cit. on p. 454).

– (1972). "Some topics in differential algebraic geometry. I. Analytic subgroups of algebraic groups". *Amer. J. Math.* 94, pp. 1195–1204. MR: 0435088 (cit. on p. 456).

Alan Baker (1975). *Transcendental number theory*. Cambridge University Press, London-New York, pp. x+147. MR: 0422171 (cit. on p. 454).

Benjamin Bakker and Jacob Tsimerman (Dec. 2017). "The Ax-Schanuel conjecture for variations of Hodge structures". arXiv: 1712.05088 (cit. on pp. 458, 469).

D. Bertrand and W. Zudilin (2003). "On the transcendence degree of the differential field generated by Siegel modular forms". *J. Reine Angew. Math.* 554, pp. 47–68. MR: 1952168.

Manjul Bhargava, Arul Shankar, Takashi Taniguchi, Frank Thorne, Jacob Tsimerman, and Yongqiang Zhao (Jan. 2017). "Bounds on 2-torsion in class groups of number fields and integral points on elliptic curves". arXiv: 1701.02458 (cit. on p. 466).

E. Bombieri, D. Masser, and U. Zannier (1999). "Intersecting a curve with algebraic subgroups of multiplicative groups". *Internat. Math. Res. Notices* 20, pp. 1119–1140. MR: 1728021 (cit. on pp. 468, 469).

E. Bombieri and J. Pila (1989). "The number of integral points on arcs and ovals". *Duke Math. J.* 59.2, pp. 337–357. MR: 1016893 (cit. on p. 460).

James A. Carlson and Domingo Toledo (2014). "Compact quotients of non-classical domains are not Kähler". In: *Hodge theory, complex geometry, and representation theory*. Vol. 608. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 51–57. MR: 3205511 (cit. on p. 458).

Pierre Colmez (1993). "Périodes des variétés abéliennes à multiplication complexe". *Ann. of Math. (2)* 138.3, pp. 625–683. MR: 1247996 (cit. on p. 467).

Pierre Deligne (1971). "Travaux de Shimura", 123–165. Lecture Notes in Math., Vol. 244. MR: 0498581.

– (1979). "Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques". In: *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*. Proc. Sympos. Pure Math., XXXIII. Amer. Math. Soc., Providence, R.I., pp. 247–289. MR: 546620.

Lou van den Dries (1998). *Tame topology and o-minimal structures*. Vol. 248. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, pp. x+180. MR: 1633348 (cit. on p. 459).

Lou van den Dries, Angus Macintyre, and David Marker (1994). "The elementary theory of restricted analytic fields with exponentiation". *Ann. of Math. (2)* 140.1, pp. 183–205. MR: 1289495.

Lou van den Dries and Chris Miller (1994a). "On the real exponential field with restricted analytic functions". *Israel J. Math.* 85.1-3, pp. 19–56. MR: 1264338.

– (1994b). "On the real exponential field with restricted analytic functions". *Israel J. Math.* 85.1-3, pp. 19–56. MR: 1264338 (cit. on p. 459).

– (1996). "Geometric categories and o-minimal structures". *Duke Math. J.* 84.2, pp. 497–540. MR: 1404337.

Bas Edixhoven (2005). "Special points on products of modular curves". *Duke Math. J.* 126.2, pp. 325–348. MR: 2115260 (cit. on p. 465).

Jordan S. Ellenberg and Akshay Venkatesh (2007). "Reflection principles and bounds for class group torsion". *Int. Math. Res. Not. IMRN* 1, Art. ID rnm002, 18. MR: 2331900 (cit. on p. 466).

A. M. Gabrièlov (1968). "Projections of semianalytic sets". *Funkcional. Anal. i Priložen.* 2.4, pp. 18–30. MR: 0245831 (cit. on p. 459).

Ziyang Gao (2017). "Towards the Andre-Oort conjecture for mixed Shimura varieties: the Ax-Lindemann theorem and lower bounds for Galois orbits of special points". *J. Reine Angew. Math.* 732, pp. 85–146. MR: 3717089 (cit. on pp. 457, 462).

P. Habegger and J. Pila (2012). "Some unlikely intersections beyond André-Oort". *Compos. Math.* 148.1, pp. 1–27. MR: 2881307.

Jun-Muk Hwang and Wing-Keung To (2002). "Volumes of complex analytic subvarieties of Hermitian symmetric spaces". *Amer. J. Math.* 124.6, pp. 1221–1246. MR: 1939785 (cit. on p. 463).

David Kazhdan (1983). "On arithmetic varieties. II". *Israel J. Math.* 44.2, pp. 139–159. MR: 693357.

A. G. Khovanskiĭ (1991). *Fewnomials*. Vol. 88. Translations of Mathematical Monographs. Translated from the Russian by Smilka Zdravkovska. American Mathematical Society, Providence, RI, pp. viii+139. MR: 1108621 (cit. on p. 459).

Jonathan Kirby (2009). "The theory of the exponential differential equations of semiabelian varieties". *Selecta Math. (N.S.)* 15.3, pp. 445–486. MR: 2551190.

B. Klingler (n.d.). "Hodge Loci and Atypical Intersections: Conjectures" (cit. on p. 458).

B. Klingler, E. Ullmo, and A. Yafaev (2016). "The hyperbolic Ax-Lindemann-Weierstrass conjecture". *Publ. Math. Inst. Hautes Études Sci.* 123, pp. 333–360. MR: 3502100 (cit. on pp. 457, 462).

Bruno Klingler and Andrei Yafaev (2014). "The André-Oort conjecture". *Ann. of Math. (2)* 180.3, pp. 867–925. MR: 3245009.

Serge Lang (1966). *Introduction to transcendental numbers*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., pp. vi+105. MR: 0214547 (cit. on p. 464).

Stanisław Łojasiewicz (1991). *Introduction to complex analytic geometry*. Translated from the Polish by Maciej Klimek. Birkhäuser Verlag, Basel, pp. xiv+523. MR: 1131081.

D. W. Masser (1977). "Division fields of elliptic functions". *Bull. London Math. Soc.* 9.1, pp. 49–53. MR: 0444572 (cit. on p. 466).

David Masser and Gisbert Wüstholz (1993). "Isogeny estimates for abelian varieties, and finiteness theorems". *Ann. of Math. (2)* 137.3, pp. 459–472. MR: 1217345 (cit. on p. 467).

J. Milne (n.d.). "Introduction to Shimura varieties" (cit. on p. 456).

J. S. Milne (1983a). "Kazhdan's theorem on arithmetic varieties".

– (1983b). "The action of an automorphism of **C** on a Shimura variety and its special points". In: *Arithmetic and geometry, Vol. I*. Vol. 35. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 239–265. MR: 717596.

N. Mok (n.d.). "Zariski closures of images of algebraic subsets under the uniformization map of finite-volume quotients of the complex unit ball" (cit. on p. 464).

Ngaiming Mok, Jonathan Pila, and Jacob Tsimerman (Nov. 2017). "Ax-Schanuel for Shimura varieties". arXiv: 1711.02189 (cit. on pp. 457, 458, 461).

Ben Moonen (1998). "Linearity properties of Shimura varieties. I". *J. Algebraic Geom.* 7.3, pp. 539–567. MR: 1618140 (cit. on p. 456).

Martin Orr (Oct. 2017). "Unlikely intersections with Hecke translates of a special subvariety". arXiv: 1710.04092 (cit. on p. 469).

Ya'acov Peterzil and Sergei Starchenko (2004). "Uniform definability of the Weierstrass $\wp$ functions and generalized tori of dimension one". *Selecta Math. (N.S.)* 10.4, pp. 525–550. MR: 2134454.

– (2008). "Complex analytic geometry in a nonstandard setting". In: *Model theory with applications to algebra and analysis. Vol. 1*. Vol. 349. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 117–165. MR: 2441378.

– (2009). "Complex analytic geometry and analytic-geometric categories". *J. Reine Angew. Math.* 626, pp. 39–74. MR: 2492989.

– (2010). "Tame complex analysis and o-minimality". In: *Proceedings of the International Congress of Mathematicians. Volume II*. Hindustan Book Agency, New Delhi, pp. 58–81. MR: 2827785 (cit. on p. 461).

– (2013). "Definability of restricted theta functions and families of abelian varieties". *Duke Math. J.* 162.4, pp. 731–765. MR: 3039679 (cit. on p. 462).

J. Pila and A. J. Wilkie (2006). "The rational points of a definable set". *Duke Math. J.* 133.3, pp. 591–616. MR: 2228464 (cit. on p. 460).

Jonathan Pila (2011). "O-minimality and the André-Oort conjecture for $\mathbb{C}^n$". *Ann. of Math. (2)* 173.3, pp. 1779–1840. MR: 2800724 (cit. on p. 457).

Jonathan Pila and Jacob Tsimerman (2013). "The André-Oort conjecture for the moduli space of abelian surfaces". *Compos. Math.* 149.2, pp. 204–216. MR: 3020307.

– (2014). "Ax-Lindemann for $\mathcal{A}_g$". *Ann. of Math. (2)* 179.2, pp. 659–681. MR: 3152943 (cit. on p. 457).

– (2016). "Ax-Schanuel for the j-function". *Duke Math. J.* 165.13, pp. 2587–2605. MR: 3546969 (cit. on p. 457).

Jonathan Pila and Umberto Zannier (2008). "Rational points in periodic analytic sets and the Manin-Mumford conjecture". *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* 19.2, pp. 149–162. MR: 2411018 (cit. on pp. 465, 468).

M. Raynaud (1983a). "Courbes sur une variété abélienne et points de torsion". *Invent. Math.* 71.1, pp. 207–233. MR: 688265 (cit. on pp. 464, 465).

– (1983b). "Sous-variétés d'une variété abélienne et points de torsion". In: *Arithmetic and geometry, Vol. I*. Vol. 35. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 327–352. MR: 717600 (cit. on pp. 464, 465).

J.-P. Rolin, P. Speissegger, and A. J. Wilkie (2003). "Quasianalytic Denjoy-Carleman classes and o-minimality". *J. Amer. Math. Soc.* 16.4, pp. 751–777. MR: 1992825 (cit. on p. 460).

A. Seidenberg (1958). "Abstract differential algebra and the analytic case". *Proc. Amer. Math. Soc.* 9, pp. 159–164. MR: 0093655.

– (1969). "Abstract differential algebra and the analytic case. II". *Proc. Amer. Math. Soc.* 23, pp. 689–691. MR: 0248122.

Jih Min Shyr (1977). "On some class number relations of algebraic tori". *Michigan Math. J.* 24.3, pp. 365–377. MR: 0491596 (cit. on p. 466).

Jacob Tsimerman (2012). "Brauer-Siegel for arithmetic tori and lower bounds for Galois orbits of special points". *J. Amer. Math. Soc.* 25.4, pp. 1091–1117. MR: 2947946 (cit. on p. 466).

– (2015). "Ax-Schanuel and o-minimality". In: *O-minimality and diophantine geometry*. Vol. 421. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 216–221. MR: 3496448 (cit. on pp. 455, 464, 468).

– (2018). "The André-Oort conjecture for $\mathfrak{A}_g$". *Ann. of Math. (2)* 187.2, pp. 379–390. MR: 3744855 (cit. on p. 467).

Emmanuel Ullmo and Andrei Yafaev (2014). "Hyperbolic Ax-Lindemann theorem in the cocompact case". *Duke Math. J.* 163.2, pp. 433–463. MR: 3161318 (cit. on pp. 457, 465).

– (2015). "Nombre de classes des tores de multiplication complexe et bornes inférieures pour les orbites galoisiennes de points spéciaux". *Bull. Soc. Math. France* 143.1, pp. 197–228. MR: 3323347 (cit. on p. 466).

Claire Voisin (2002). *Hodge theory and complex algebraic geometry. I*. Vol. 76. Cambridge Studies in Advanced Mathematics. Translated from the French original by Leila Schneps. Cambridge University Press, Cambridge, pp. x+322. MR: 1967689 (cit. on p. 457).

A. J. Wilkie (1996). "Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function". *J. Amer. Math. Soc.* 9.4, pp. 1051–1094. MR: 1398816 (cit. on p. 459).

Masaaki Yoshida (1987). *Fuchsian differential equations*. Aspects of Mathematics, E11. With special emphasis on the Gauss-Schwarz theory. Friedr. Vieweg & Sohn, Braunschweig, pp. xiv+215. MR: 986252.

Xinyi Yuan and Shou-Wu Zhang (2018). "On the averaged Colmez conjecture". *Ann. of Math. (2)* 187.2, pp. 533–638. arXiv: 1507.06903. MR: 3744857 (cit. on p. 467).

Boris Zilber (2002). "Exponential sums equations and the Schanuel conjecture". *J. London Math. Soc. (2)* 65.1, pp. 27–44. MR: 1875133 (cit. on p. 468).

Received 2018-02-28.

Jacob Tsimerman
Department of Mathematics, University of Toronto, Toronto, Canada
jacobt@math.toronto.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (473–484)

SHARP SPHERE PACKINGS

Maryna Viazovska

Abstract

In this talk we will speak about recent progress on the sphere packing problem. The packing problem can be formulated for a wide class of metric spaces equipped with a measure. An interesting feature of this optimization problem is that a slight change of parameters (such as the dimension of the space or radius of the spheres) can dramatically change the properties of optimal configurations. We will focus on those cases when the solution of the packing problem is particularly simple. Namely, we say that a packing problem is sharp if its density attains the so-called linear programming bound. Several such configurations have been known for a long time and we have recently proved that the E_8 lattice sphere packing in $\mathbb{R}^8$ and the Leech lattice packing in $\mathbb{R}^{24}$ are sharp. Moreover, we will discuss common unusual properties of shared by such configurations and outline possible applications to Fourier analysis.

1 Introduction

The classical sphere packing problem asks for the densest possible configuration of non-overlapping equal balls in the three dimensional Euclidean space. This natural and even naive question remained open for several centuries and has driven a lot of research in geometry, combinatorics and optimization. The complete proof of the sphere packing problem was given by T. Hales in 1998 Hales [2005].

A similar question can be asked for Euclidean spaces of dimensions other then three or for spaces with other geometries, such as a sphere, a projective space, or the Hamming space. The packing problem is not only an exciting mathematical puzzle, it also plays a role in computer science and signal processing as a mathematical model of the error correcting codes.

In this paper we will focus on the upper bounds for the sphere packing densities. There exist different methods for proving such bounds. One conceptually simple and still rather

MSC2010: 94B65.

powerful approach is the linear programming. We are particularly interested in those packing problems, which can be completely solved by this method. We will call such arrangements of balls the *sharp packings*.

The sharp packings have many interesting properties. In particular, the distribution of pairwise distances between the centers of sharply packed spheres gives rise to summation and interpolation formulas. In the last section of this paper we will discuss a new interpolation formula for the Schwartz functions on the real line.

2 Linear programming bounds for sphere packings in metric spaces

Let (M, dist) be a metric space equipped with a measure μ. For $x \in M$ and $r > 0$ we denote by $B(x,r)$ the open ball with center x and radius r. Let X be a discrete subset of M such that $\mathrm{dist}(x,y) \geq 2r$ for any distinct $x, y \in X$. Then the set $\mathcal{P} := \cup_{x \in X} B(x,r)$ is a *sphere packing* in M. We define the *density* of $\mathcal{P}$ as

$$\Delta_{\mathcal{P}} := \sup_{x_0 \in M} \limsup_{R \to \infty} \frac{\mu(\mathcal{P} \cap B(x_0, R))}{\mu(B(x_0, R))}.$$

Our goal is to search for densest possible configurations and to prove upper bounds on the packing density.

The *linear programming* is a powerful and simple method to prove upper bounds for the packing problems. This technique was successfully applied to obtain upper bounds in a wide range of discrete optimization problems such as error-correcting codes Delsarte [1972], equal weight quadrature formulas Delsarte, Goethals, and Seidel [1977], and spherical codes Kabatiansky and Levenshtein [1978] and Pfender and Ziegler [2004]. In this section we explain the idea behind this method, consider several examples, and discuss the limitations of this approach.

The essence of the linear programming method is the replacement of a complicated geometrical optimization problem by a simpler convex optimization problem.

We say that a function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is *geometrically positive* (with respect to a metric space M) if

$$\sum_{x,y \in Y} g(\mathrm{dist}(x,y)) \geq 0$$

for all finite subsets $Y \subset M$.

We can obtain an upper bound for the packing density by solving the following convex optimization problem. For simplicity we assume that M is compact.

Lemma 2.1. *Fix $r > 0$. Let $g_r : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a function and c_0 be a positive constant such that*

(i) $g_r - c_0$ is geometrically positive
(ii) $g_r(t) \le 0$ for all $t \in [2r, \infty)$. Then any packing of balls of radius r in M has cardinality at most

$$\frac{g_r(0)}{c_0}.$$

Proof. Let $X \subset M$ be a subset such that $\operatorname{dist}(x, y) \ge 2r$ for any pair of distinct points $x, y \in X$. Then condition (i) implies

$$\sum_{x,y \in X} g_r(\operatorname{dist}(x, y)) = \sum_{x,y \in X} (g_r(\operatorname{dist}(\mathrm{x}, \mathrm{y})) - c_0) + |X|^2 c_0 \ge |X|^2 c_0.$$

On the other hand, by condition (ii)

$$\sum_{x,y \in X} g_r(\operatorname{dist}(x, y)) = |X| g_r(0) + \sum_{\substack{x,y \in X \\ x \ne y}} g_r(\operatorname{dist}(x, y)) \le |X| g_r(0).$$

Hence, we arrive at

$$|X| \le \frac{g_r(0)}{c_0}.$$

□

Unfortunately, the description of the cone of geometrically positive functions is usually a very difficult problem. Therefore, we will consider a smaller cone, the cone of so-called positive-definite functions. A function $p : \mathbb{R}_{\ge 0} \to \mathbb{R}$ is *positive definite* (with respect to a metric space M) if

$$\sum_{x,y \in Y} w_x \, w_y \, p(\operatorname{dist}(x, y)) \ge 0$$

for all finite subsets $Y \subset M$ and all collections of real weights $\{w_y\}_{y \in Y}$. For metric spaces M with a big isometry group the cone of positive definite functions has a simple description in terms of representation theory.

Theorem 2.2. *(Bocher 1941) Let G be a topological group acting continuously on a topological space M. For every G-invariant positive-definite kernel $p : M \times M \to \mathbb{C}$, there exists a unitary representation V of G and a continuous, G-equivariant map $\phi : M \to V$ such that $p(x, y) = \langle \phi(x), \phi(y) \rangle$ for all $x, y \in M$.*

For example, the following theorem characterizes positive definite functions on the standard sphere $S^{d-1} = \{x \in \mathbb{R}^d \mid \|x\|^2 = 1\}$. Let $P_k^d(t)$ denote the degree k ultraspherical (i.e. Gegenbauer) polynomial, normalized with $P_k^d(1) = 1$. These polynomials are orthogonal with respect to the measure $(1 - t^2)^{(d-3)/2} \, dt$ on $[-1, 1]$.

Theorem 2.3. *(Schoenberg [1942]) A function $g : [0,2] \to \mathbb{R}$ is positive definite with respect to the sphere S^{d-1} if and only if*

$$g(s) = \sum_{k=0}^{\infty} c_k P_k^d(1 - \frac{1}{2}s^2)$$

where $c_k \geq 0$.

H. Cohn and N. Elkies have applied the linear programming technique to the sphere packing problem in Euclidean space Cohn and Elkies [2003]. This problem is rather subtle since the Euclidean space in non-compact and the Lebesgue measure of the whole space is not finite.

Let us setup some notations in order to formulate the main result of Cohn and Elkies [ibid.]. The *Fourier transform* of an L^1 function $f : \mathbb{R}^d \to \mathbb{C}$ is defined as

$$\mathfrak{F}(f)(y) = \widehat{f}(y) := \int\limits_{\mathbb{R}^d} f(x)\, e^{-2\pi i x\cdot y}\, dx, \quad y \in \mathbb{R}^d$$

where $x \cdot y = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x - y\|^2$ is the standard scalar product in $\mathbb{R}^d$. A C^∞ function $f : \mathbb{R}^d \to \mathbb{C}$ is called a *Schwartz function* if it tends to zero as $\|x\| \to \infty$ faster then any inverse power of $\|x\|$, and the same holds for all partial derivatives of f. The set of all Schwartz functions is called the *Schwartz space*. The Fourier transform is an automorphism of this space. We will also need the following wider class of functions. We say that a function $f : \mathbb{R}^d \to \mathbb{C}$ is *admissible* if there is a constant $\delta > 0$ such that $|f(x)|$ and $|\widehat{f}(x)|$ are bounded above by a constant times $(1 + |x|)^{-d-\delta}$. The following theorem is the key result of Cohn and Elkies [ibid.]:

Theorem 2.4 (Cohn and Elkies [ibid.]). *Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is an admissible function, $r_0 \in \mathbb{R}_{>0}$ and they satisfy:*

(2-1) $$f(x) \leq 0 \text{ for } \|x\| \geq r_0,$$

(2-2) $$\widehat{f}(x) \geq 0 \text{ for all } x \in \mathbb{R}^d$$

and

(2-3) $$f(0) = \widehat{f}(0) = 1.$$

Then the density of d-dimensional sphere packings is bounded above by

$$\frac{\pi^{\frac{d}{2}} r_0^d}{2^d\, \Gamma(\frac{d}{2} + 1)} = \frac{f(0)}{\widehat{f}(0)} \cdot \operatorname{Vol} B_d(0, \frac{r_0}{2}).$$

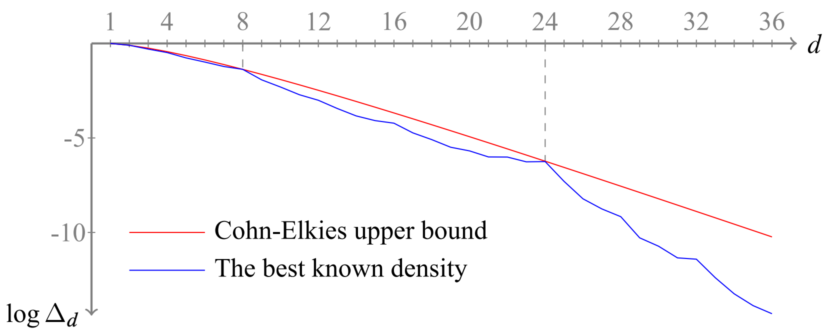


Figure 1: Upper and lower bounds for Δ_d

The *sphere packing constant* Δ_d is the supremum of all densities of sphere packings in $\mathbb{R}^d$.

Cohn and Elkies have numerically applied Theorem 2.4 to the sphere packing constant in dimensions from 1 to 36. The numerical results obtained in Cohn and Elkies [ibid.] are illustrated in Figure 1. The red line represents an upper bound obtained from Theorem 2.4 and the blue line shows the density of the best known configuration in each dimension.

3 Sharp linear programming bounds

A natural question is whether the linear programming bounds can be sharp. As we have relaxed our original optimization problem, we do not expect sharp bounds in general. However, we know several examples when the linear programming technique provides a complete solution to the optimization problem.

A beautiful example is the computation of the kissing number in dimensions 8 and 24. We recall, that the *kissing number* $K(d)$ is the maximal number of "blue" spheres that can touch a "red" sphere of the same size in d-dimensional Euclidean space. It was proven by Odlyzko and Sloane [1979] and independently by Levenshtein [1979] that $K(8) = 240$ and $K(24) = 196560$. The proof of this result is based on the linear programming method. Let us consider the kissing problem in dimension 8 in more detail. The kissing configuration can be described as follows. Consider 112 vectors of type $(0^6, \pm 2^2)$ that is, with 2 non-zero coordinates, which are ± 2 and 128 vectors of type $(\pm 1^8)$ with an even number of positive components. All the $122 + 128 = 240$ vectors have length $2\sqrt{2}$. The minimum distance between these vectors also equals $2\sqrt{2}$. Therefore, they form a kissing configuration. The only missing step is a construction of a suitable positive definite function p_8. Consider the following polynomial on $[-1,1]$:

$$p_8(t) := (t+1)^2\,(t+\frac{1}{2})^2\,t^2\,(t-\frac{1}{2}).$$

The coefficients c_k of the expansion of p_8 in Gegenbauer polynomials are all non-negative and $p_8(1)/c_0 = 240$.

The dimensions 8 and 24 are also special for the sphere packing problem in the Euclidean space. On the Figure 1 we can see that the blue line representing a lower bound on the sphere packing constant and the red line representing Cohn-Elkies bound come very close together at the dimensions 8 and 24. In Cohn and Elkies [2003] Cohn and Elkies proved the following estimates

Theorem 3.1 (Cohn and Elkies [ibid.]). *We have*

$$\Delta_8 \leq 1.00016\,\Delta_{E_8},$$

$$\Delta_{24} \leq 1.019\,\Delta_{\Lambda_{24}}.$$

Here Δ_{E_8} denotes the density of the E_8-lattice packing in $\mathbb{R}^8$ and $\Delta_{\Lambda_{24}}$ denotes the density of the Leech lattice packing in $\mathbb{R}^{24}$.

It is proven in M. S. Viazovska [2017] and Cohn, Kumar, Miller, Radchenko, and M. Viazovska [2017] that the Cohn-Elkies linear programming bounds are indeed sharp in these dimensions. The sphere packing problem in dimensions 8 and 24 will be discussed in more detail in the next section.

At the end of this section, we would like to mention several packing problems for which the numerical linear programming bounds are extremely close to the known lower bounds, however the question whether these bounds are sharp is still open. The first example, is the packing of equal disks in dimension 2. The packing problem itself has been solved long time ago Thue [1910], Fejes [1943] by a geometric method. The numerical results of Cohn and Elkies [2003] suggest that the linear programming bound is also sharp in this case, however the exact solution is not known yet.

There is a numerical evidence that the packing problem can be solved by linear programming also for other convex center symmetric bodies in $\mathbb{R}^2$. H. Cohn and G. Minton have numerically studied the packings with translates 2-dimensional of L_p-balls using linear programming bounds proven in Cohn and Elkies [ibid., Theorem B.1]. Recall, that for $p > 0$ a p-ball in $\mathbb{R}^d$ is the set of points $x = (x_1, \ldots, x_d)$ such that

$$|x_1|^p + \ldots + |x_d|^p \leq 1.$$

Cohn and Minton conjecture that the resulting bounds are sharp. Thanks to a theorem proven by L. Fejes Tóth we know that the optimal packing of congruent convex center symmetric bodies in $\mathbb{R}^2$ is always a lattice packing. An open question is whether this result can be proven by linear programming.

Finally, interesting numerical results has been obtained for translative packings of L_p-balls in $\mathbb{R}^3$. On Figure 2 we plot the upper and lower bounds for such packings computed in Dostert [2017]. We know that the classical sphere packing problem in dimension 3 can not be solved by linear programming. However, for L_p-balls with parameter p in the interval $(1.2, 1.4)$ the lower and upper bounds come extremely close. So there is a hope that these bounds are sharp for some values of p.

4 The sphere packing problem in dimensions 8 and 24

In this section we will consider the sphere packing problem in the Euclidean spaces of dimensions 8 and 24.

It the 8-dimensional Euclidean space there exists a highly structured configuration – the E_8 lattice, which we have already mentioned in Section 3. The E_8-lattice $\Lambda_8 \subset \mathbb{R}^8$ is given by

$$\Lambda_8 = \{(x_i) \in \mathbb{Z}^8 \cup (\mathbb{Z} + \tfrac{1}{2})^8 | \ \textstyle\sum_{i=1}^8 x_i \equiv 0 \pmod 2\}.$$

Λ_8 is the unique even, unimodular lattice of rank 8. The minimal distance between two points in Λ_8 is $\sqrt{2}$. The E_8-lattice sphere packing is the packing of unit balls with centers at $\sqrt{2}\Lambda_8$.

The following theorem implies that the optimality of the E_8-lattice sphere packing can be proven by the Cohn-Elkies method.

Theorem 4.1. *(M. S. Viazovska [2017]) There exists a radial Schwartz function $f_{E_8} : \mathbb{R}^8 \to \mathbb{R}$ which satisfies:*

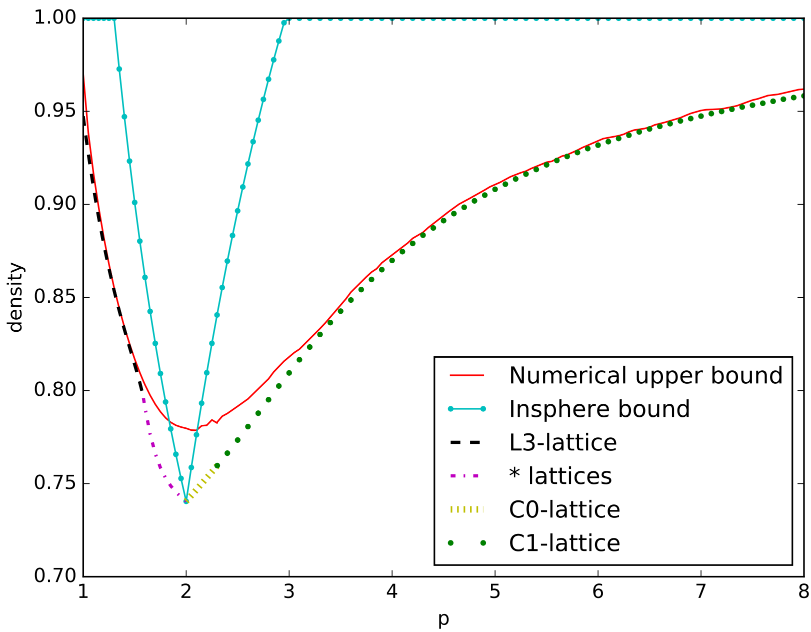
$$f_{E_8}(x) \le 0 \text{ for } \|x\| \ge \sqrt{2}$$
$$\widehat{f}_{E_8}(x) \ge 0 \text{ for all } x \in \mathbb{R}^8$$
$$f_{E_8}(0) = \widehat{f}_{E_8}(0) = 1.$$

An immediate corollary of Theorems 2.4 and 4.1.

Theorem 4.2. *No packing of unit balls in Euclidean space $\mathbb{R}^8$ has density greater than that of the E_8 lattice packing. Therefore $\Delta_8 = \frac{\pi^4}{384} \approx 0.25367$.*

Also in dimension 24 there exists a lattice with unusually tight structure. The Leech lattice Λ_{24} is an even, unimodular lattice of rank 24. The minimal distance between two points in Λ_{24} is 2, and it is the only even, unimodular lattice of rank 24 with this property. The Leech lattice sphere packing is the packing of unit balls with centers at Λ_{24}. The

Figure 2: Lower and upper bounds for the density of translative packings of p-balls in $\mathbb{R}^3$ computed in Dostert [2017].



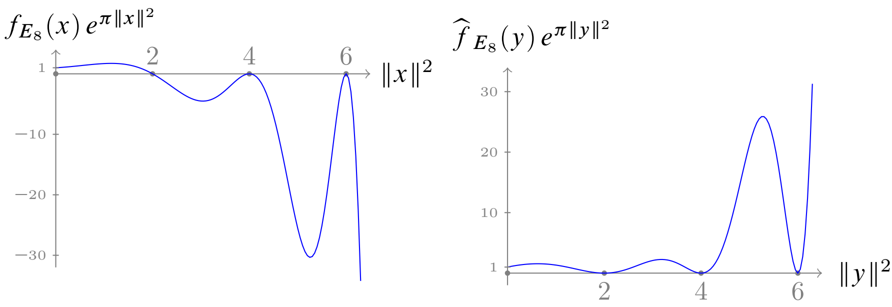


Figure 3: Plot of the functions f_{E_8} and $\widehat{f}_{E_8}$

optimality of this packing also has been proved by the Cohn-Elkies linear programming method.

Theorem 4.3. *(Cohn, Kumar, Miller, Radchenko, and M. Viazovska [2017]) There exists a radial Schwartz function $f_{\Lambda_{24}} : \mathbb{R}^{24} \to \mathbb{R}$ which satisfies:*

$$f_{\Lambda_{24}}(x) \leq 0 \textit{ for } \|x\| \geq 2$$
$$\widehat{f}_{\Lambda_{24}}(x) \geq 0 \textit{ for all } x \in \mathbb{R}^{24}$$
$$f_{\Lambda_{24}}(0) = \widehat{f}_{\Lambda_{24}}(0) = 1.$$

This result immediately implies

Theorem 4.4. *(Cohn, Kumar, Miller, Radchenko, and M. Viazovska [ibid.]) No packing of unit balls in the Euclidean space $\mathbb{R}^{24}$ has density greater than that of the Leech lattice packing. Therefore $\Delta_{24} = \frac{\pi^{12}}{12!} \approx 0.00193$.*

Remarks:

1. Without loss of generality we may assume that f_{E_8} is radial.

2. By the Poisson summation formula we have

$$f_{E_8}(0) \geq \sum_{\ell \in \Lambda_8} f_{E_8}(\ell) = \sum_{\ell \in \Lambda_8} \widehat{f}_{E_8}(\ell) \geq \widehat{f}_{E_8}(0).$$

This can happen only if $f_{E_8}(\sqrt{2n}) = \widehat{f}_{E_8}(\sqrt{2n}) = 0$ for all $n \in \mathbb{Z}_{>0}$.

5 Fourier interpolation

The idea behind our construction of f_{E_8} and $f_{\Lambda_{24}}$ is the hypothesis that a radial Schwartz function p can be uniquely reconstructed from the values

$$\{p(\sqrt{2n}),\, p'(\sqrt{2n}),\, \widehat{p}(\sqrt{2n}),\, \widehat{p}'(\sqrt{2n})\}_{n=0}^{\infty}$$

The proof of this statement is a goal an ongoing project of the author in collaboration with H. Cohn, A. Kumar, S. D. Miller, and D. Radchenko.

In this section we will present a simpler first degree interpolation formula of this type.

Theorem 5.1. *(Radchenko, Viazovska Radchenko and M. Viazovska [2017]) There exists a collection of Schwartz functions $b_0, a_n\colon \mathbb{R}\to\mathbb{R}$ with the property that for any Schwartz function $p\colon \mathbb{R}\to\mathbb{R}$ and any $x\in\mathbb{R}$ we have*

$$\begin{aligned} p(x) =&c_0(x)\, p'(0) + \sum_{n\in\mathbb{Z}} a_n(x)\, p(\operatorname{sign}(n)\sqrt{|n|}) \\ &+ \widehat{c}_0(x)\, p'(0) + \sum_{n\in\mathbb{Z}} \widehat{a}_n(x)\widehat{p}(\operatorname{sign}(n)\sqrt{|n|}), \end{aligned} \tag{5-1}$$

where the right-hand side converges absolutely.

Moreover, we can describe all possible collections of values of a Schwartz function at the points $\{\pm\sqrt{n}\}_{n=0}^{\infty}$.

Denote by $\mathfrak{s}$ the vector space of all rapidly decaying sequences of real numbers, i.e., sequences $(x_n)_{n\geq 0}$ such that for all $k>0$ we have $n^k x_n\to 0, n\to\infty$.

We denote by $\mathcal{S}$ the space of Schwartz functions on $\mathbb{R}$. Consider the map $\Psi\colon \mathcal{S}\to \mathbb{R}^2\oplus\mathfrak{s}\oplus\mathfrak{s}$ given by

$$\Psi(p) = \left(p'(0), \widehat{p}'(0), (p(\operatorname{sign}(n)\sqrt{|n|}))_{n\in\mathbb{Z}}, (\widehat{p}(\operatorname{sign}(n)\sqrt{|n|}))_{n\in\mathbb{Z}}\right).$$

Theorem 5.2. *(Radchenko, Viazovska Radchenko and M. Viazovska [ibid.]) The map Ψ is an isomorphism between the space of Schwartz functions and the vector space* $\ker L\subset \mathbb{R}^2\oplus\mathfrak{s}\oplus\mathfrak{s}$, *where $L\colon \mathbb{R}^2\oplus\mathfrak{s}\oplus\mathfrak{s}\to\mathbb{R}^2$ is the linear map*

$$\begin{aligned} L :&(x_0', y_0', (x_n)_{n\in\mathbb{Z}}, (y_n)_{n\in\mathbb{Z}}) \mapsto \\ &\Bigg(\sum_{n\in\mathbb{Z}} x_{n^2} - \sum_{n\in\mathbb{Z}} y_{n^2}, \\ &2x_0' + \sum_{n\in\mathbb{Z}} \operatorname{sign}(n)\frac{r_3(|n|)\,x_n}{\sqrt{|n|}} - 2iy_0' - \sum_{n\in\mathbb{Z}} i\operatorname{sign}(n)\frac{r_3(|n|)y_n}{\sqrt{|n|}}\Bigg). \end{aligned}$$

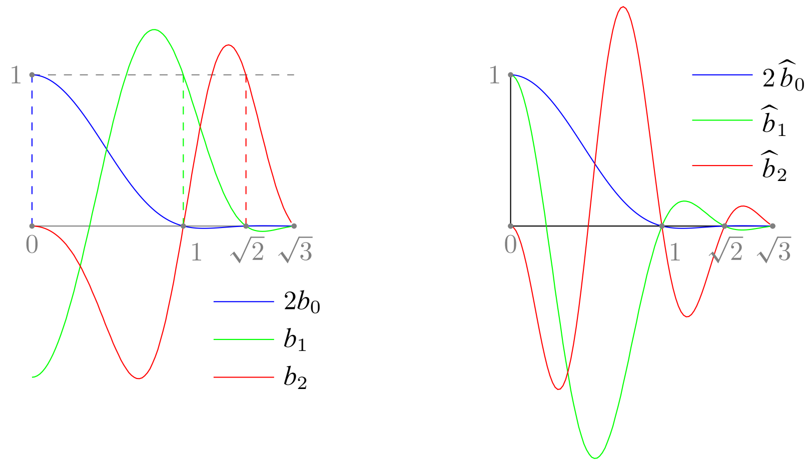


Figure 4: Plots of $b_n(x) := a_n(x) + a_n(-x)$ and $\widehat{b}_n$ for $n = 0, 1, 2$.

Also Theorem 5.1 allows us to construct an unusual family of discrete measures on the real line. A *cristalline measure* on $\mathbb{R}^d$ is a tempered distribution μ such that μ and $\widehat{\mu}$ are both charges with locally finite support. A simplest example of a cristalline measure is the *Dirac comb*

$$\mu_{\text{Dirac}} = \sum_{n\in\mathbb{Z}} \delta_n.$$

Recently, Lev and Olevskii [2015] have proven that crystalline measures with uniformally discrete support and spectrum (the support of the Fourier transform) can be obtained from the Dirac comb by dilations, shifts, multiplication on exponentials, and taking linear combinations.

Theorem 5.3. *(Lev and Olevskii [ibid.]) Let μ be a crystalline measure on $\mathbb{R}$ with uniformally discrete support and spectrum. Then the support of μ is contained in a finite union of translates of a certain lattice L. Moreover, μ is of the form*

$$\mu = \sum_{j=1}^{N} P_j \sum_{\lambda\in L+\theta_j} \delta_\lambda$$

where θ_j, $j = 1, \ldots, N$ are real numbers and P_j, $j = 1, \ldots, N$ are trigonometric polynomials.

The interpolation formula implies that there exists a continuous family of *exotic* crystalline measures

$$\mu_x := \delta_x + \delta_{-x} - \sum_{n=0}^{\infty} b_n(x)\,(\delta_{\sqrt{n}} + \delta_{-\sqrt{n}}).$$

Let us briefly explain our strategy for the construction of the interpolating basis a_n, c_0 introduced in Theorem 5.1. We will separately consider the odd and even components of the Schwartz functions. We set $b_n(x) = a_n(x) + a_n(-x)$. Then the symmetry implies $b_n = b_{-n}$.

Let us consider the generating series formed by the functions $\{b_n\}_{n=0}^{\infty}$ and their Fourier transforms. For $x \in \mathbb{R}$ and a complex number τ with $\Im(\tau) > 0$ we define

$$F(x,\tau) := \sum_{n=0}^{\infty} b_n(x)\, e^{\pi i n \tau}$$

$$\widetilde{F}(x,\tau) := \sum_{n=0}^{\infty} \widehat{b}_n(x)\, e^{\pi i n \tau}.$$

We will show that these two functions satisfy a functional equation with respect to the variable τ. Indeed, the interpolation formula interpolation formula (5-1) applied to the Gaussian $e^{\pi i x^2 \tau}$ gives

$$e^{\pi i x^2 \tau} = F(x,\tau) + \frac{1}{\sqrt{-i\tau}}\widetilde{F}(x, \frac{-1}{\tau}).$$

In Radchenko and M. Viazovska [2017] we solve this functional equation using the theory of modular integrals. A similar idea also leads to the constuction of functions f_{E_8} and $f_{\Lambda_{24}}$ in Theorems 4.1 and 4.3.

References

Henry Cohn and Noam Elkies (2003). "New upper bounds on sphere packings. I". *Ann. of Math. (2)* 157.2, pp. 689–714. MR: 1973059 (cit. on pp. 476–478).

Henry Cohn and Abhinav Kumar (2007). "Universally optimal distribution of points on spheres". *J. Amer. Math. Soc.* 20.1, pp. 99–148. MR: 2257398.

Henry Cohn, Abhinav Kumar, Stephen D. Miller, Danylo Radchenko, and Maryna Viazovska (2017). "The sphere packing problem in dimension 24". *Ann. of Math. (2)* 185.3, pp. 1017–1033. MR: 3664817 (cit. on pp. 478, 481).

J. H. Conway and N. J. A. Sloane (1995). "What are all the best sphere packings in low dimensions?" *Discrete Comput. Geom.* 13.3-4, pp. 383–403. MR: 1318784.

P. Delsarte (1972). "Bounds for unrestricted codes, by linear programming". *Philips Res. Rep.* 27, pp. 272–289. MR: 0314545 (cit. on p. 474).

P. Delsarte, J. M. Goethals, and J. J. Seidel (1977). "Spherical codes and designs". *Geometriae Dedicata* 6.3, pp. 363–388. MR: 0485471 (cit. on p. 474).

M. Dostert (2017). "Geometric Packings of Non-Spherical Shapes". PhD. thesis. PhD thesis. Universität zu Köln (cit. on pp. 479, 480).

L. Fejes (1943). "Über die dichteste Kugellagerung". *Math. Z.* 48, pp. 676–684. MR: 0009129 (cit. on p. 478).

Thomas C. Hales (2005). "A proof of the Kepler conjecture". *Ann. of Math. (2)* 162.3, pp. 1065–1185. MR: 2179728 (cit. on p. 473).

G. A. Kabatiansky and V. I. Levenshtein (1978). "Bounds for packings on a sphere and in space". *Problems of Information Transmission* 14, pp. 1–17 (cit. on p. 474).

Nir Lev and Alexander Olevskii (2015). "Quasicrystals and Poisson's summation formula". *Invent. Math.* 200.2, pp. 585–606. MR: 3338010 (cit. on p. 483).

V. I. Levenshtein (1979). "On bounds for packings in n-dimensional euclidean space". *Soviet Math. Doklady* 20, pp. 417–421 (cit. on p. 477).

A. M. Odlyzko and N. J. A. Sloane (1979). "New bounds on the number of unit spheres that can touch a unit sphere in n dimensions". *J. Combin. Theory Ser. A* 26.2, pp. 210–214. MR: 530296 (cit. on p. 477).

Florian Pfender and Günter M. Ziegler (2004). "Kissing numbers, sphere packings, and some unexpected proofs". *Notices Amer. Math. Soc.* 51.8, pp. 873–883. MR: 2145821 (cit. on p. 474).

Danylo Radchenko and Maryna Viazovska (Jan. 2017). "Fourier interpolation on the real line". arXiv: 1701.00265 (cit. on pp. 482, 484).

I. J. Schoenberg (1942). "Positive definite functions on spheres". *Duke Math. J.* 9, pp. 96–108. MR: 0005922 (cit. on p. 476).

A. Thue (1910). "Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene". *Norske Vid. Selsk. Skr.* 1, pp. 1–9 (cit. on p. 478).

Maryna S. Viazovska (2017). "The sphere packing problem in dimension 8". *Ann. of Math. (2)* 185.3, pp. 991–1015. MR: 3664816 (cit. on pp. 478, 479).

Received 2018-02-27.

MARYNA VIAZOVSKA
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
1025 LAUSANNE
SWITZERLAND
viazovska@gmail.com

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (485–504)

CHARACTERISTIC SUBSETS AND THE POLYNOMIAL METHOD

Miguel N. Walsh

Abstract

We provide an informal discussion of the polynomial method. This is a tool of general applicability that can be used to exploit the algebraic structure arising in some problems of arithmetic nature.

1 Introduction

1.1 The polynomial method. This article provides an informal discussion of the polynomial method. For us, this is the idea that when studying a problem with an underlying algebraic structure, a set S may admit a simpler "characteristic subset" $A \subseteq S$ that controls S, in the sense that a polynomial vanishing at A with sufficiently high multiplicity is forced to vanish at all or most points of S. By using dimension counting arguments in the spirit of Siegel's lemma, one may then exploit the simplicity of A to find a polynomial with suitable characteristics vanishing at most points of S.

This idea has been applied in a wide variety of contexts. Our choice of topics is largely based on personal taste, but we do try to convey the multitude of areas where it is relevant, the similarities in how the method is applied in them and some connections that exist between the different subjects. A variant of the method where the dimension counting arguments are used to find a polynomial that produces an adequate partition of the given points, instead of vanishing at them, has proven remarkably useful in recent work and is also treated in this article. We do not present proofs, but many are discussed. We refer the reader to Guth [2016c] and Tao [2014] for some further surveys on this circle of ideas.

Part of this article was written while the author was a Clay Research Fellow.
MSC2010: primary 11B30; secondary 52C10, 42B10, 11L07, 14G05, 37A45.

1.2 Notation. Before we proceed let us summarise the notation that will be used. Given two quantities X and Y, we will write either $X = O(Y)$ or $X \lesssim Y$ to mean that there exists some absolute constant C such that the inequality $X \leq CY$ holds. If this constant depends on some other parameter d, we may indicate this using a subscript and write $X = O_d(Y)$ or $X \lesssim_d Y$. If A is a finite set of points, $|A|$ will stand for its cardinality. We shall write $\mathbb{F}_q$ for the finite field with q elements, while the notation $\mathbb{F}$ alone is meant to stand for an arbitrary field. Finally, given a polynomial $P \in \mathbb{F}[x_1, \ldots, x_n]$, we shall write $Z(P) = \{x \in \mathbb{F}^n : P(x) = 0\}$ for its zero set.

2 Incidence problems

2.1 Incidence geometry and its applications. We shall choose incidence geometry as our starting point and therefore spend some time motivating this subject. Some of the questions incidence geometry is concerned with rank among the simplest questions one can formulate in mathematics, but to define its problems broadly we consider the following set-up. Let V be an algebraic variety over a field $\mathbb{F}$, T a finite family of subvarieties of V and S a finite set of points inside of V. Incidence geometry is then mainly concerned with how the quantity

$$I(S,T) = |\{(s,t) \in S \times T : s \in t\}|,$$

which counts the number of incidences between S and T, relates to the sizes of S and T.

For example, when the variety V above is equal to $\mathbb{R}^n$ and T is just an arbitrary finite family of lines, we may ask for the maximal number of incidences that can occur between a set of points and a set of lines. A classical result of Szemerédi and Trotter [1983] gives in this case the sharp asymptotic bound

$$I(S,T) \lesssim |S|^{2/3}|T|^{2/3} + |S| + |T|,$$

where the implicit constant is uniform among all choices of S and T. If, more generally, we ask what happens when T is a finite family of algebraic curves of degree at most d, then a corresponding bound of the form

$$I(S,T) \lesssim_d |S|^{\frac{d^2+1}{2d^2+1}}|T|^{\frac{2d^2}{2d^2+1}} + |S| + |T|,$$

was established by Pach and Sharir [1998].

These two results constitute very simple instances of the general context described before. There is a vast literature dealing with many different cases that may arise, with T ranging from families of circles to high-dimensional varieties and with $\mathbb{F}$ ranging from a finite field to the complex numbers. While in principle there is no reason why some general unifying results could not be established, they certainly seem hard to come by. Similarly, obtaining sharp asymptotic bounds that avoid extra factors depending on $|S|$ or $|T|$ tends to be an additional challenge. Even when $V = \mathbb{R}^n$ and each element of T is a hypersurface defined by a single irreducible polynomial of bounded degree this question is not fully settled.

The simplicity of its questions is arguably one of the things that makes incidence geometry attractive and it is also this simplicity what makes its results find natural applications in different contexts. To give an example of this, let us discuss a very simple and direct connection to arithmetic combinatorics. To a finite set of points A in $\mathbb{R}^n$, say, we can associate the sets

$$A + A = \{a + a' : a, a' \in A\},$$

and

$$A \cdot A = \{a \cdot a' : a, a' \in A\}.$$

It is expected that both sets cannot be small simultaneously Erdős and Szemerédi [1983] and an estimate of this form was established by Elekes [1997] using incidence geometry. His idea was to consider the set of all lines of the form $y = (x - a)a'$, with $a, a' \in A$. Clearly, when $x = b + a$ for some $b \in A$, we get that $y = a'b \in A \cdot A$. As a consequence, we see that each of these lines touches $|A|$ points of the grid $(A+A)\times(A\cdot A)$. If both $A+A$ and $A \cdot A$ were to be small, this would then mean that these lines are highly incident to each other. So much so in fact, that they would contradict the Szemerédi-Trotter theorem. Thus we get the desired result.

Incidence bounds in the broader context we have described at the beginning of this section give rise to more general results concerning the size of

$$f(A_1, \ldots, A_m) = \{f(a_1, \ldots, a_m) : a_1 \in A_1, \ldots, a_m \in A_m\},$$

where f is a polynomial and $A_1, \ldots, A_m$ are finite sets of points (see for example Elekes and Szabó [2012]). Results of this type can in particular be applied to obtain randomness extractors, opening the door for incidence geometry to be applied in theoretical computer science Dvir [2010]. It should also be noted that the above sum-product phenomenon, as it is known, and the related concept of expansion, are pervasive throughout different parts of mathematics and have found a number of remarkable applications. See, for example, this survey by Helfgott [2015] and Green's 2014 ICM article Green [2014] for some discussion.

Sometimes problems can be encapsulated as incidence questions in more subtle ways. An example is provided by the Erdős distinct distances conjecture Erdős [1946], which asked for a bound on the minimal number of pairwise distances determined by n distinct points in the plane. Here Elekes and Sharir [2010] started with the simple observation that two pairs of points at the same distance determine two segments of the same length and so there exists some rigid motion taking one segment to the other. As it happens, given two points $x, y \in \mathbb{R}^2$, the set of rigid motions that take x to y can essentially be viewed as a line in $\mathbb{R}^3$ under an appropriate parametrisation of most of $\mathrm{SE}(2)$ (the special Euclidean group in two-dimensions). As a consequence, two pairs x_1, x_2 and y_1, y_2 at the same distance correspond to an intersection between the line of motions that send x_1 to x_2 and the line of motions that send y_1 to y_2. In paritcular, n points determining few distances lead to a set of highly incident lines in $\mathbb{R}^3$. Continuing this sort of arguments and exploiting some additional properties of the resulting set of lines, they managed to reduce Erdős' conjecture to an incidence question that was subsequently settled by Guth and Katz [2015].

Finally, the nature of incidence geometry also makes it a good source of toy models for more difficult problems. One notorious example is given by the following well-known conjecture in geometric measure theory.

Conjecture 2.1 (Kakeya problem)**.** *A set $E \subseteq R^n$ containing a unit line segment in every direction must have Hausdorff dimension equal to n.*

A counterexample to this problem would certainly be reminiscent of the existence of a set of lines being highly incident to each other while satisfying some restricting conditions regarding their directions. This problem served as a motivation for a large body of work in incidence geometry and we will now see how it led to the introduction of the polynomial method in this area.

2.2 Enter the polynomial method. There a number of ways of formulating toy models for the Kakeya problem in incidence geometry. A particularly straightforward way of doing so is to formulate the problem over $\mathbb{F}_q$, where the discrete nature of this space naturally turns it into a question about incidences. The resulting Kakeya problem over finite fields was answered in the affirmative by Dvir [2009]. Precisely, he established the following result.

Theorem 2.2 (Kakeya over finite fields)**.** *Let $K \subseteq \mathbb{F}_q^n$ be a set containing a line in every direction. Then $|K| \gtrsim q^n$.*

This question was originally posed by Wolff [1999] and attracted considerable attention. Despite the large amount of work that preceded his article, Dvir's proof is extremely simple. What distinguishes it is its introduction of the polynomial method as the main tool to attack the problem. While this tool had barely been used in this or related areas before, the situation would change dramatically after this result.

The polynomial method is largely concerned with finidng a polynomial with suitable properties vanishing at most points of a set of interest. To find such a polynomial, an important role is played by the following simple observation, usually attributed to Thue and Siegel, which allows us to bound the degree of a polynomial vanishing on an arbitrary set.

Lemma 2.3 (Siegel's lemma). *For every set finite $S \subseteq \mathbb{F}^n$ there exists a non-zero polynomial $P \in \mathbb{F}[x_1, \ldots, x_n]$, of degree $\lesssim_n |S|^{1/n}$, vanishing on S.*

The proof is extremely simple, consisting only of a dimension counting argument. There are so many different polynomials of degree $\lesssim_n |S|^{1/n}$ that at least two of them, say P_1 and P_2, must take the same values on S. Therefore $P = P_1 - P_2$ is of the form we want. The same kind of argument can be used to force some additional properties on the polynomial P, like having integer coefficients or vanishing to high multiplicity on our set of points.

To obtain a polynomial of low degree vanishing on a given set S, the polynomial method combines the above lemma with what might be called the idea of characteristic subsets (here we are using the notation of M. N. Walsh [2012b]). We informally define these as follows.

Definition 2.4 (Characteristic subset). A set $A \subseteq S$ is said to be a characteristic subset of S if any polynomial of low complextiy that vanishes with a certain multiplicity on A must also vanish at all or most points of S.

What exactly does it mean for a polynomial P to have low complexity depends very much on what is needed in the given problem, though it always implies that the degree of P should be small. In some cases, we may also require the coefficients to be small or restricted to the integers, or both. Clearly, if a set S admits a characteristic subset A of small size, then we can find a polynomial of low degree vanishing at all or most points of S by applying Siegel's lemma to A. As we will see during this article, this idea is a central part of the polynomial method.

Let us now see how it applies to the proof of Theorem 2.2. Dvir's crucial observation is that any set $K \subseteq \mathbb{F}_q^n$ containing a line in every direction must be a characteristic subset of $\mathbb{F}_q^n$. This is easiest to see looking at the corresponding projective space. If P is a polynomial of degree strictly less than q vanishing on an affine line in every direction, by Bezout's theorem its homogenisation must also contain the hyperplane at infinity in its zero set and this necessarily implies that P vanishes at the whole space. As a consequence, if we could find a nontrivial polynomial of degree less than q that vanishes on K, it would also have to vanish on $\mathbb{F}_q^n$, which is impossible. Comparing this observation with Siegel's lemma, we conclude that it must be $|K| \gtrsim_n q^n$.

2.3 Polynomial partitioning.

The idea of characteristic subsets will also play a role in incidence geometry results over $\mathbb{R}^n$. For example, if we are given a highly incidence family of lines T, we may be able to find a subset of lines $T' \subseteq T$ such that any polynomial of low degree P vanishing on all the elements of T' must also vanish on most elements of T. The reason why this may work is that, given the highly incident nature of T, we may be able to find a small subset $T' \subseteq T$ such that most elements $t \in T$ will be incident to many lines in T'. As a consequence, a polynomial P vanishing on T' will vanish at many points of such t. If the degree of P is sufficiently small, Bezout's theorem would then force P to vanish on all of t as desired. This kind of approach is often called degree-reduction in this context.

Although arguments like this tend to be useful, it turns out that a substantial addition to the polynomial method is required in order to make progress on incidence problems taking place in $\mathbb{R}^n$. Since the original work of Szemerédi and Trotter, a central idea in incidence geometry has been to partition the space into cells in a way that limits how many of these cells the elements of T may intersect. This plays into the intuitive notion that the varieties being studied in an incidence problem should spread out across space, thus forbidding them to cluster into a highly incident configuration. As it turns out, polynomials can be used to obtain such a partition of space in an extremely structured way. Indeed, the following result was establish by Guth and Katz [2015] when trying to adapt Dvir's polynomial method to incidence questions over Euclidean space.

Theorem 2.5 (Polynomial partitioning of $\mathbb{R}^n$). *For every finite set $S \in \mathbb{R}^n$ and every choice of an integer $M \geq 1$, there exists some nonzero polynomial $P \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $\lesssim_n M$ such that each connected component of $\mathbb{R}^n \setminus Z(P)$ contains at most $\lesssim_n \frac{|S|}{M^n}$ points of S.*

By an old result of Petrovskiĭ and Oleĭnik [1949] it is known that if P is a polynomial in n variables of degree $O(M)$, then $\mathbb{R}^n \setminus Z(P)$ can have at most $O(M^n)$ connected components. Theorem 2.5 then says that at any level M of our choice and for any set S, we can partition the points of S as efficiently as possible among the connected components of the complement of a polynomial of degree at most $O(M)$. The generality of such a statement is quite striking. There is one caveat, however, which is that the result does not rule out the possibility that some of the points of S actually lie inside of $Z(P)$. But this in itself may be an advantage, since we may be able to exploit the additional structure of having a proper subvariety covering our set of points in order to attack the problem that we are interested in.

Let us now briefly discuss how a partition of S of the above type can be used to tackle questions in incidence geometry. For simplicity, let us assume we are studying the incidences of S with a set of lines T and let us apply Theorem 2.5 to S for an adequate choice of M. Then we obtain a partition by the zero set of some polynomial P of degree $O(M)$ in such a way that each connected component of $\mathbb{R}^n \setminus Z(P)$ contains very few points of S. But since a line that is not properly contained in $Z(P)$ can touch this zero set in at most $\deg(P)$ places, it follows that such a line can only intersect $\deg(P)+1 = O(M)$ of the components of $\mathbb{R}^n \setminus Z(P)$. In other words, each component contains few points and each line intersects few components. Combining this observation with a trivial incidence estimate in each component coming from the Cauhcy-Schwarz inequality, this readily provides a sharp incidence bound upon making an adequate choice of the parameter M.

It then remains to deal with the incidences occurring inside of $Z(P)$. This points out why the general context described at the beginning of this section can be relevant even if one is only interested in estimates over $\mathbb{R}^n$. To obtain sharp estimates in this case, the study of $Z(P)$ sometimes requires one to pay attention to more subtle algebraic properties of the subvarieties involved (see for example Katz's ICM 2014 article Katz [2014]), but many times even this can be avoided if one is willing to settle for slightly weaker bounds.

In any case, the fact that the problem is reduced to the study of subvarieties is not capricious. To see this, consider for example the task of estimating the number of incidences between a set of points and a set of lines in $\mathbb{R}^3$. The first part of the argument above that deals with the incidences occurring inside the cells, when applied in $\mathbb{R}^3$, leads to a bound that improves the one provided by the Szemerédi-Trotter theorem. However, the latter result is sharp in $\mathbb{R}^2$ and so it is clear that this improvement is not possible in general. Indeed, an example showing that the Szemerédi-Trotter theorem is optimal in $\mathbb{R}^2$ can obviously be replicated inside any plane of $\mathbb{R}^3$. This is a general phenomenon: incidence bounds can be worse than expected if the elements of T cluster inside lower-dimensional varieties. In their work, Guth and Katz managed to obtain an improved bound in $\mathbb{R}^3$ under the assumption that such a clustering fails to take place. Part of the remarkable effectiveness of the polynomial method as discussed above is that it provides optimal bounds outside of the zero set of a certain polynomial and thus singles out the correct obstruction, that is, the possibility that many incidences are occurring inside lower-dimensional varieties.

3 Number theory

3.1 Stepanov's method. We will now depart from the incidence geometry questions discussed in the previous section and see how the polynomial method makes an appearance throughout a range of topics in number theory. We will see how it was recently used in the study of the distribution of sets in residue classes mod p and to bound the number of rational points on curves. However, the polynomial method in this context is not new and was used by Stepanov [1969], and later Bombieri [1974], in the related topic of estimating the number of $\mathbb{F}_q$-points on a curve. Precisely, they provided an alternative proof of the following estimate, equivalent to the Riemann Hypothesis for curves over finite fields.

Theorem 3.1 (Hasse-Weil bound)**.** *Let $\mathcal{C}$ be a nonsingular absolutely irreducible projective curve of genus g defined over $\mathbb{F}_q$. Then, writing $\mathcal{C}(q)$ for the number of $\mathbb{F}_q$-points of*

$\mathcal{C}$*, we have the estimate*

$$|\mathcal{C}(q) - (q+1)| \leq 2g\sqrt{q}.$$

The story of the Riemann Hypothesis for curves over finite fields is well-known. It was originally conjectured by Artin [1924], with the case of elliptic curves being established by Hasse [1936]. The general result was famously obtained by Weil [1949] in work that laid the foundation of modern algebraic geometry. However, an alternative proof that is more elementary was subsequently found by Stepanov in special cases using what we would now call the polynomial method. His method was subsequently used by Bombieri to produce a very simple argument that handles the result in full generality.

Let us now briefly discuss how the proof works. The lower bound on $\mathcal{C}(q)$ provided by Theorem 3.1 can be easily derived from the upper bound by means of a lifting trick, so we will just discuss how the latter is established. We are interested in an upper bound on the number of $\mathbb{F}_q$-points lying inside a curve $\mathcal{C}$, or in other words, the number of points x inside of this curve that are invariant under the Frobenius map $\mathrm{Frob}(x) = x^q$. By looking at the cartesian product $\mathcal{C} \times \mathcal{C}$, we see that it will suffice to provide an upper bound for the size of the intersection of two curves in this cartesian product: the curve $\mathcal{C}_1 = \{(x, y) \in \mathcal{C} \times \mathcal{C} : x = y\}$ and the curve $\mathcal{C}_2 = \{(x, y) \in \mathcal{C} \times \mathcal{C} : \mathrm{Frob}(x) = y\}$.

A naive application of Bezout's theorem to bound the size of this intersection would fail to produce the kind of bound we want. On the other hand, Bezout's theorem does tell us that given an irreducible curve of bounded degree, a polynomial of low degree vanishing with high multiplicity on a large subset of this curve must vanish on the whole curve. Applying this to our problem we conclude that if $\mathcal{C}_1 \cap \mathcal{C}_2$ is large, then it must be a characteristic subset of $\mathcal{C}_2$. Thus in order to prove Theorem 3.1 it would suffice to construct a polynomial of low degree vanishing with high multiplicity on $\mathcal{C}_1$, but not vanishing on $\mathcal{C}_2$. This in turn follows from a combination of the Riemann-Roch theorem and the same kind of dimension counting arguments used in the proof of Siegel's lemma.

3.2 The inverse sieve problem. One of the main topics in analytic number theory is the study of the distribution of sets in residue classes mod p. Many times, the goal is to show that a special set, like the primes, is essentially equidistributed among these classes. It is then natural to wonder what kind of structure may cause a set to be badly distributed in residue classes and in particular, whether an inverse theorem may be obtained characterising all sets exhibiting bad behaviour.

As we have just seen, algebraic curves constitute one such example and in general, so do algebraic sets of higher dimension. That abnormal behaviour should always be attributable to the presence of algebraic structure was suggested by a number of authors. The following result established a conjecture of Helfgott and Venkatesh [2009] to this effect.

Theorem 3.2 (M. N. Walsh [2014]). *Let $S \subseteq \{1, \dots, N\}^d$ occupy $\ll p^\kappa$ residue classes for every prime p and some real number $0 \le \kappa < d$. Then, for every $\varepsilon > 0$, there exists some nonzero $P \in \mathbb{Z}[x_1, \dots, x_d]$ of degree $\ll_{\kappa,d,\varepsilon} (\log N)^{\frac{\kappa}{d-\kappa}}$ vanishing on at least $(1-\varepsilon)|S|$ points of S.*

The assumption that the set occupies few classes is not prohibitive and can be replaced by an estimate on its L^2-norm mod p. Furthermore, it was shown in M. N. Walsh [2012b] that the polynomial P can be taken to have degree $O(1)$ as long as the set S satisfies some regularity assumptions. In fact, these assumptions are automatically met if the set S is not small, as conjectured in Helfgott and Venkatesh [2009].

These results are essentially sharp and their proof employs the polynomial method. The argument goes more or less as follows. If S occupies few residue classes mod p for many primes p, it should be possible to find a very small set $A \subseteq S$ that, for many primes p, contains a representative of many of the classes occupied by S mod p. We claim that A is then a characteristic subset of S. Indeed, suppose P is a polynomial with small coefficients and small degree that vanishes on A. By construction of A, we find that to most elements $s \in S$ we can associate many primes p_i such that $s \equiv x_i \pmod{p_i}$ for some $x_i \in A$ and so, in particular, $P(s) \equiv P(x_i) \pmod{p_i}$. The fact that P vanishes on A then implies that every such p_i must divide $P(s)$. But since P has small coefficients and small degree, $|P(s)|$ is small, and so the only way this can happen is if $P(s) = 0$. We have thus shown that A is indeed a characteristic subset of S. The result then follows from applying Siegel's lemma to find a polynomial of low degree vanishing on A.

There is an interesting question that remains when considering one dimensional sets $S \subseteq \{1, \dots, N\}$. We know by the large sieve inequality that a set occupying approximately half of the residue classes mod p, for all primes p, can have size at most $O(N^{1/2})$. On the other hand, the squares in $\{1, \dots, N\}$ show that this estimate is sharp. The question arises whether every set of comparable size occupying at most half the residue classes mod p, for every prime p, must be correlated to the set of squares. This question can be generalised a bit further (see M. N. Walsh [2012b]). Given the arguments described above, one would expect that the polynomial method should be useful to make progress on this problem, although this has not been achieved so far (but see Green and Harper [2014] and Hanson [2017] for some partial progress by means of different tools).

3.3 The determinant method. The polynomial method has also been used effectively to bound the number of points S of bounded height that an algebraic variety can have over $\mathbb{Z}$ or $\mathbb{Q}$. As in Stepanov's method, the problem does not lie in finding an adequate characteristic subset A for S, since in this context this is generally accomplished by simply picking a maximal algebraically independent subset of S. The actual difficulty lies instead in finding an appropriate polynomial vanishing on A. This gives rise to the study of a certain system of linear equations and in particular, to estimates on the size of the determinants associated with this system. As a consequence, this form of the polynomial method has been known as the determinant method in this context. It dates back to the work of Bombieri and Pila [1989], with subsequent improvements of this method being obtained by Heath-Brown [2002] and Salberger [2010], among others.

For the rest of this discussion let us assume for simplicity that A lies inside a plane curve $\mathcal{C}$ defined by an irreducible polynomial f. The task of finding a polynomial vanishing on A may seem circular, since after all, we already know that f vanishes on A. The idea is instead to show that the dimension of the space of polynomials of small degree vanishing on A is big. Big enough, in fact, as to guarantee the existence of at least one polynomial g in this space that is not divisible by f. Since S must then lie in $\mathcal{C} \cap Z(g)$, a bound for its size readily follows from Bezout's theorem.

While the Stepanov-Bombieri argument required us at this stage to improve on Siegel's lemma by means of a dimension counting argument relying on the Riemann-Roch theorem, the improvement of Siegel's lemma needed here can be obtained through the following estimate of Bombieri and Vaaler on the space of solutions of a system of linear equations over the integers.

Theorem 3.3. *Bombieri and Vaaler [1983] Let $\sum_{k=1}^{r} b_{mk} x_k = 0$, $m = 1, \dots, s$, be a system of s linearly independent equations in $r > s$ unknowns, with integer coefficients b_{mk}. Then, there exists a nontrivial integer solution $(x_1, \dots, x_r)$ satisfying the bound*

$$\max_{1 \le i \le r} |x_i| \le \left(D^{-1} \sqrt{\left|\det\left(BB^T\right)\right|} \right)^{\frac{1}{r-s}}. \tag{3-1}$$

Here $B = (b_{mk})$ is the $s \times r$ matrix of coefficients, B^T its transpose, and D is the greatest common divisor of the determinants of the $s \times s$ minors of B.

When trying to find a polynomial of small degree that vanishes on A, the system of equations we are interested in is the one where the coefficients b_{mk} are given by the values taken by the monomials of small degree when evaluated at the elements of A. It is the presence of the factor D^{-1} in the above statement what provides an improvement over the classical form of Siegel's lemma.

We already know that as a consequence of the Hasse-Weil bound, the rows of the resulting matrix of coefficients B will occupy few residue classes mod p for many primes p, exceptions occurring only when $\mathcal{C}$ has a singular reduction mod p. On the other hand, the discussion of the previous subsection would suggest that a system of linear equations whose coefficients occupy few residue classes mod p, for many primes p, should be easier to solve. In a sense, Theorem 3.3 formalises this idea. Indeed, if the rows of B occupy few classes mod p then we would expect its minors to be divisible by a high power of p. A concrete estimate of this form is established in Salberger's work, giving rise to a strong lower bound on the size of D. This lower bound, when combined with Theorem 3.3, forces A to be small and as a consequence, the dimension of the space of polynomials vanishing on A to be large, as desired.

Let us finally remark that it is an interesting feature of the polynomial method that when the points being studied are already known to lie in the zero set of a polynomial with either large coefficients or large degree, this can usually be used to obtain stronger estimates. As noted by Ellenberg and Venkatesh [2005], something like this holds true in this context. In particular, it can be used to produce an improved lower bound for the left-hand side of (3-1), by assuming for the sake of contradiction that all solutions are multiples of f. This in turn gives rise to improved bounds for S when f has a large coefficient and in particular, manages to compensate a loss that appears in Salberger's bound in this case. Combining the ideas we have discussed, a uniform bound of the form $O_d(N^{2/d})$ was obtained in M. N. Walsh [2015] for the number of rational points of height at most N that an irreducible curve of degree d can have. It is easy to see that this bound is asymptotically sharp upon consideration of the equation $y = x^d$.

3.4 Exponential sums and Montgomery's conjecture. We finish this section with a brief discussion of some connections between the topics studied so far. On the one hand, the problems treated in the previous section are naturally related with each other, with such incidence problems ultimately leading to the Kakeya problem over $\mathbb{R}^n$ and the more general Stein's restriction conjecture in harmonic analysis. As we shall see in Section 4.4, the polynomial method can be extended to make progress on these problems as well.

On the other hand, we have seen that the number-theoretic topics discussed in this section are also quite interconnected. Furthermore, estimates like the large sieve inequality and the Riemann Hypothesis for curves over finite fields lead into the general area of exponential sums, where some of the most far-reaching conjectures in analytic number theory have been formulated. An outstanding example is the exponent pairs conjecture Iwaniec and Kowalski [2004], an open problem that has among its consequences the Lindelöf Hypothesis. This part of mathematics gives us an excuse to join the two lines of enquiry we have covered so far in this article. In particular, not only the density hypothesis for the zeros of the Riemann zeta function but also the Kakeya problem would follow from a positive answer to the following conjecture about exponential sums.

Conjecture 3.4 (Montgomery's Conjecture Montgomery [1971]). *For any real number $r \geq 1$ and any sequence of complex numbers $(a_n)_{n=1}^N$ with $|a_n| \leq 1$, the estimate*

$$\frac{1}{T}\int_0^T \left|\sum_{n=1}^{N} a_n n^{is}\right|^{2r} ds \lesssim_\varepsilon N^{r+\varepsilon},$$

holds for all $T \geq N^r$ and all $\varepsilon > 0$.

That this implies the Kakeya problem can be seen upon rephrasing the latter as a question about the existence of small subsets of $\mathbb{F}_p$ containing large arithmetic progressions with each possible common difference Bourgain [1991]. There is indeed a good amount of work making progress on the Kakeya problem employing tools from arithmetic combinatorics.

While this shows that analytic number theory could be used to make progress on problems related to the Kakeya problem, the opposite is also true, with progress on questions related to restriction theory being used recently to yield results in analytic number theory. Indeed, the l^2-decoupling theory started by Bourgain and Demeter [2015], a family of results in the spirit of the restriction conjecture whose proof relies on the multilinear Kakeya inequality among other things, has been used to establish Vinogradov's main conjecture in analytic number theory Bourgain, Demeter, and Guth [2016]. Previous to this work, the best result on this problem had been obtained by Wooley [2012] by means of his efficient congruencing mod p method. Finally, the set of ideas surrounding the decoupling theory was also used by Bourgain to improve the best-known exponent towards the Lindelöf Hypothesis Bourgain [2017].

It is then fair to ask to what extent the polynomial method is a tool that finds applications on a wide variety of contexts and to what extent it reflects underlying phenomenona in somewhat interconnected families of results.

4 Further topics

4.1 Baker's theorem. In order to emphasise the recurring features of the polynomial method, let us briefly discuss one last example of an application that would seem to have little connection to the topics discussed in the rest of this article, besides its link to arithmetic. Our choice is Baker's classical result in transcendental number theory regarding linear forms in logarithms Baker [1968]. For simplicity, let us discuss the integer case, where we are seeking uniform lower bounds over expressions of the form $\sum_{i=1}^{m} b_i \log a_i$, where the a_i are multiplicatively independent integers and the b_i are integers not all equal to zero.

To attack this problem, we will consider the curve $\mathcal{C} = (a_1^t, \dots, a_m^t)$, and more generally, subsets of this curve of the form

$$\mathcal{C}_N = \{(a_1^n, \dots, a_m^n) : n = 1, \dots, N\},$$

for positive integers N. Looking at the corresponding Vandermonde matrix of coefficients, the fact that the a_i are multiplicatively independent easily implies that no polynomial P of small degree, relative to N, can vanish at all points of $\mathcal{C}_N$.

A polynomial P evaluated at the curve $\mathcal{C} = (a_1^t, \dots, a_m^t)$ may be seen as a function of the parameter t. With this point of view, we claim that we can find some value N_0, relatively small with respect to N, such that $\mathcal{C}_{N_0}$ is a characteristic subset of $\mathcal{C}_N$. Indeed, suppose we are given some polynomial P with integer coefficients and low degree that vanishes at $\mathcal{C}_{N_0}$ with some large multiplicity J with respect to the variable t. Then, $\mathcal{C}$ being an analytic curve, this will force $\frac{d^j}{dt^j} P$ to take small values in a neighbourhood of $\mathcal{C}_{N_0}$, as long as j is at least slightly smaller than J. In particular, this will happen at all points of $\mathcal{C}_{N_1}$, provided N_1 is an integer not much larger than N_0. If J is sufficiently large, we can then iterate this argument enough times as to guarantee that P itself takes small values at all points of $\mathcal{C}_N$. But since P takes integer values, the only way this can happen is if P is in fact zero at all points of $\mathcal{C}_N$.

Since $\mathcal{C}_{N_0}$ is a characteristic subset of $\mathcal{C}_N$, we know, by our observation regarding the Vandermonde matrix, that no polynomial with integer coefficients and low degree can vanish with high multiplicity at $\mathcal{C}_{N_0}$. On the other hand, if $\sum_{i=1}^m b_i \log a_i$ were to be small, an argument like in the proof of Siegel's lemma can be used to contradict this fact. Indeed, for any polynomial P with integer coefficients, a relation of this kind between the $\log a_i$ significantly restricts the range of values that the derivatives of P with respect to t can take at any given point of $\mathcal{C}_{N_0}$. In particular, this allows us to find two polynomials P_1, P_2 with integer coefficients and abnormally low degree such that, with respect to t, their first J derivatives take the same values at $\mathcal{C}_{N_0}$, for some large value of J. The polynomial $P = P_1 - P_2$ then gives us the desired contradiction.

The generality of Baker's result makes it applicable in a number of different contexts, thus implicitly extending the range of problems where the polynomial method may have some relevance. As a curious example, the integer version just discussed was used by Bourgain, Lindenstrauss, Michel, and Venkatesh [2009] to provide effective proofs of a family of results relating to Furstenberg's $\times 2 \times 3$ conjecture in ergodic theory Furstenberg [1967]. These include the well-known Rudolph-Johnson theorem Rudolph [1990] and Johnson [1992], establishing the conjecture in the positive entropy case, and Furstenberg's topological result, showing that $\mathbb{R}/\mathbb{Z}$ has no infinite closed subset other than itself that is invariant under multiplication by a pair of multiplicatively independent integers Furstenberg [1967]. One may wonder whether proofs that use the polynomial method in an explicit way may contribute to understand better such applications of Baker's theorem.

4.2 Structure and randomness. It may be worthwhile to give some brief consideration to how the polynomial method fits with the general phenomenon of structure-randomness decompositions that is pervasive throughout analysis. The idea of the latter is that it tends to be possible to decompose an object of interest into a part that detects the structure of the problem being studied and a random part that is, in a sense, completely independent from this information. In the context of Hilbert spaces, this observation can be formalised by noting that given a distinguished set Σ of bounded functions, any bounded element of the Hilbert space may be written in the form $\sum_j \lambda_j \sigma_j + g$, where the sum of the coefficients λ_j is uniformly bounded, $\sigma_j \in \Sigma$ for every j, and g is essentially orthogonal to Σ, in the sense that $\langle g, \sigma \rangle$ is small for every $\sigma \in \Sigma$. We refer the reader to this article Gowers [2010] of Gowers for an elegant discussion of this and more general decompositions.

The polynomial method may be considered a useful complement to the structure-randomness approach when the arithmetic of the problem gives rise to an underlying algebraic structure. The characteristic subsets notation used in M. N. Walsh [2012b, 2014] and in the present article is, in fact, inspired by the concept of characteristic factors that plays the role of the structured part in several problems in ergodic theory Furstenberg [1967], Furstenberg and Weiss [1996], and Host and Kra [2005]. In the same way that information about the behaviour of a polynomial on a set can be deduced from what happens in a characteristic subset, the behaviour of nonconventional ergodic averages over a set of functions can be deduced from what happens in the corresponding characteristic factors. This is exactly what these decompositions seek to accomplish.

Structure-randomness decompositions are a flexible tool that can be substantially refined when a small error term is allowed in the decompositions. This played an important role in the convergence result of M. N. Walsh [2012a] and we refer again to Gowers' article Gowers [2010] for a general discussion. As noted by Lovász and Szegedy [2007], these more general decompositions can be seen as versions of the Szemerédi regularity lemma Szemerédi [1978]. The regularity lemma was itself the key tool used to obtain the cell-decomposition in the original proof of the Szemerédi-Trotter theorem and, in this sense, the polynomial partitioning method of Guth an Katz may be seen as an instance of these general decomposition results where the algebraic nature of the decomposition is made more explicit.

4.3 Polynomial partitioning over varieties. In order to apply the polynomial method efficiently over general varieties some further improvements may be needed. During the discussion of the determinant method, we observed how the polynomial method could be made more effective when the points being studied lie in the zero set of some polynomial with large coefficients. One may similarly wonder whether knowing that the points lie inside of an algebraic variety of high degree may also lead to improved estimates. Since it is known that the dimension of the polynomial ring associated with a variety increases in proportion with its degree Nesterenko [1984], it is logical to suspect that a corresponding improvement may be achieved on the kind of dimension counting arguments that are ubiquitous in the polynomial method. Indeed, combining this kind of observations with the type of tools used to prove Theorem 2.5, we can obtain the following result.

Theorem 4.1 (Polynomial partitioning over varieties M. Walsh [n.d.]). *Given any real algebraic variety $V \subseteq \mathbb{R}^n$ of dimension d, any set of points $S \subseteq V$ and any integer $M \geq 1$, there exists some polynomial $P \in \mathbb{R}[x_1, \dots, x_n]$ of degree $\lesssim_{d,n} M$, not vanishing identically on V, such that each connected component of $V \setminus Z(P)$ contains at most $\lesssim_{d,n} \frac{|S|}{M^d \deg(V)}$ points of S.*

That this should hold was already conjectured in Basu and Sombra [2016]. A more general sharp estimate was obtained in M. Walsh [n.d.] by including an additional explicit dependence on the degrees of the various individual polynomials defining the variety V. Notice that a particular case of the above result is a version of Siegel's lemma that allows us to find a polynomial of degree $\lesssim_{d,n} \frac{|S|^{1/d}}{\deg(V)^{1/d}}$ that vanishes on S without vanishing identically on V.

As we saw in Section 2.3, both the nature of incidence geometry and of polynomial partitioning techniques lead to the consideration of lower-dimensional algebraic varieties, even for problems that originally take place over $\mathbb{R}^n$. Without proper tools to handle varieties of high degree, the polynomial partitioning needs to be truncated as to produce only varieties of low degree, leading to suboptimal bounds. Estimates like Theorem 4.1 may prove useful in providing a unified approach to manage these problems and produce sharp bounds.

Nevertheless, to make this work, we saw that a second result that is needed is a bound on how many of the components produced by the above partitioning result can be touched by a given algebraic variety W. In general, by work of Milnor [1964] and Thom [1965], we have bounds that give a good dependence in terms of the degree of the partitioning polynomial, but not in terms of W. Some progress on this issue has been made by Barone and Basu [2012], who were able to obtain an estimate with a good dependence on the product of the degrees of the polynomials defining W, with this result being subsequently applied in incidence geometry Basu and Sombra [2016].

Since in general estimates like Theorem 4.1 can only be expected to yield a saving proportional to the degree of the variety V itself, it may be necessary to obtain a version of the result of Barone and Basu that depends only on the degree of W, instead of depending on the product of the degrees of the polynomials defining it. A step in this direction was taken in M. Walsh [n.d.] where a bound with a main term of the desired form was obtained. Nevertheless, the error term in this estimate is not optimal and it remains an interesting problem to improve it. In fact, it is shown in M. Walsh [ibid.] that a suitable refinement can indeed be combined with Theorem 4.1 to attain sharp incidence bounds that currently remain out of reach.

4.4 Restriction estimates. The proof of Theorem 4.1 combines algebraic estimates for the size of ideals with tools like the polynomial ham-sandwich theorem, the latter being a result that lies at the heart of polynomial partitioning results since the original work of Guth and Katz. The classical ham-sandwich theorem states that given n open sets in $\mathbb{R}^n$, they can always be simultaneously bisected by a suitable hyperplane. The polynomial ham-sandwich theorem extends this claim to show that we can always bisect a larger number of open sets as long as, instead of restricting to hyperplanes, we allow the bisection to be performed by hypersurfaces of a correspondingly large degree.

This bisecting result can be further put to use to extend the scope of the polynomial method to the Kakeya conjecture in Euclidean space. This conjecture can be sen as a question about estimating the minimal possible volume that can be attained by a collection of tubes pointing in a large number of quantitatively distinct directions. Covering this union by a family of small cubes and replacing the notion of a polynomial vanishing at a point by that of a polynomial bisecting a cube, the ideas surrounding the polynomial method can be extended from the discrete case to this continuous setting.

Recall that to find characteristic subsets we have relied heavily on estimates on the intersection of algebraic sets. For example, that given a polynomial P not vanishing identically on a given line, this line can only intersect $Z(P)$ in at most $\deg(P)$ points. In order to carry the polynomial method to this new context, we need to find analogues of these estimates that hold true for tubes. For example, we may observe that if a polynomial P takes small values at more than $C \deg(P)$ points along a fixed tube, for some sufficiently large constant C, then it must also take small values at most places that lie between these points. Similarly, another alternative is to consider the directed volume of a surface, allowing one more or less to conclude that a polynomial P can cut a tube transversally in at most $O(\deg(P))$ places.

By introducing these ideas, Guth essentially showed in Guth [2016b] that a counterexample to the Kakeya problem in $\mathbb{R}^3$ can be approximated by a polynomial of the smallest possible degree allowed by the Crofton formula, and used this to obtain graininess estimates in the spirit of Katz, Łaba, and Tao [2000]. He also combined these ideas with slightly more sophisticated tools from algebraic topology, involving cohomology classes and Lusternik-Schnirelmann theory, to obtain the first proof Guth [2010] of the endpoint case of the multilinear Kakeya inequality of Bennett, Carbery, and Tao [2006].

It is even possible to extend the polynomial method further to make progress on the more general restriction conjecture of Stein [1979]. Here, by applying a decomposition into wave packets, we can translate the problem into a question about overlapping patterns of tubes and this can be treated in a similar spirit as the incidence geometry questions that we discussed in Section 2.3. Indeed, the polynomial partitioning method works in a very similar way, as long as we replace the zero set of the partitioning polynomial by its neighbourhood, as to be able to bound the number of cells of the resulting partition that a tube can intersect Guth [2016a]. It should be remarked that this approach helps to highlight the role played by low degree varieties in hypothetical counterexamples to this conjecture. Whether this kind of ideas based on the polynomial method will lead to further progress on this sort of problems remains an interesting question.

References

E. Artin (1924). "Quadratische Körper im Gebiete der höheren Kongruenzen. II". *Math. Z.* 19.1, pp. 207–246. MR: 1544652 (cit. on p. 492).

A. Baker (1968). "Linear forms in the logarithms of algebraic numbers. IV". *Mathematika* 15, pp. 204–216. MR: 0258756 (cit. on p. 496).

Sal Barone and Saugata Basu (2012). "Refined bounds on the number of connected components of sign conditions on a variety". *Discrete Comput. Geom.* 47.3, pp. 577–597. MR: 2891249 (cit. on p. 499).

Saugata Basu and Martín Sombra (2016). "Polynomial partitioning on varieties of codimension two and point-hypersurface incidences in four dimensions". *Discrete Comput. Geom.* 55.1, pp. 158–184. MR: 3439263 (cit. on p. 499).

Jonathan Bennett, Anthony Carbery, and Terence Tao (2006). "On the multilinear restriction and Kakeya conjectures". *Acta Math.* 196.2, pp. 261–302. MR: 2275834 (cit. on p. 500).

E. Bombieri and J. Pila (1989). "The number of integral points on arcs and ovals". *Duke Math. J.* 59.2, pp. 337–357. MR: 1016893 (cit. on p. 494).

E. Bombieri and J. Vaaler (1983). "On Siegel's lemma". *Invent. Math.* 73.1, pp. 11–32. MR: 707346 (cit. on p. 494).

Enrico Bombieri (1974). "Counting points on curves over finite fields (d'après S. A. Stepanov)", 234–241. Lecture Notes in Math., Vol. 383. MR: 0429903 (cit. on p. 491).

J. Bourgain (1991). "Remarks on Montgomery's conjectures on Dirichlet sums". In: *Geometric aspects of functional analysis (1989–90)*. Vol. 1469. Lecture Notes in Math. Springer, Berlin, pp. 153–165. MR: 1122620 (cit. on p. 496).

– (2017). "Decoupling, exponential sums and the Riemann zeta function". *J. Amer. Math. Soc.* 30.1, pp. 205–224. MR: 3556291 (cit. on p. 496).

Jean Bourgain and Ciprian Demeter (2015). "The proof of the l^2 decoupling conjecture". *Ann. of Math. (2)* 182.1, pp. 351–389. MR: 3374964 (cit. on p. 496).

Jean Bourgain, Ciprian Demeter, and Larry Guth (2016). "Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three". *Ann. of Math. (2)* 184.2, pp. 633–682. MR: 3548534 (cit. on p. 496).

Jean Bourgain, Elon Lindenstrauss, Philippe Michel, and Akshay Venkatesh (2009). "Some effective results for $\times a \times b$". *Ergodic Theory Dynam. Systems* 29.6, pp. 1705–1722. MR: 2563089 (cit. on p. 497).

Zeev Dvir (2009). "On the size of Kakeya sets in finite fields". *J. Amer. Math. Soc.* 22.4, pp. 1093–1097. MR: 2525780 (cit. on p. 488).

– (2010). "Incidence theorems and their applications". *Found. Trends Theor. Comput. Sci.* 6.4, 257–393 (2012). MR: 3004132 (cit. on p. 487).

György Elekes (1997). "On the number of sums and products". *Acta Arith.* 81.4, pp. 365–367. MR: 1472816 (cit. on p. 487).

György Elekes and Micha Sharir (2010). "Incidences in three dimensions and distinct distances in the plane [extended abstract]". In: *Computational geometry (SCG'10)*. ACM, New York, pp. 413–422. MR: 2742978 (cit. on p. 487).

György Elekes and Endre Szabó (2012). "How to find groups? (and how to use them in Erdős geometry?)" *Combinatorica* 32.5, pp. 537–571. MR: 3004808 (cit. on p. 487).

J. Ellenberg and A. Venkatesh (2005). "On uniform bounds for rational points on nonrational curves". *Int. Math. Res. Not.* 35, pp. 2163–2181. MR: 2181791 (cit. on p. 495).

P. Erdős (1946). "On sets of distances of n points". *Amer. Math. Monthly* 53, pp. 248–250. MR: 0015796 (cit. on p. 487).

P. Erdős and E. Szemerédi (1983). "On sums and products of integers". In: *Studies in pure mathematics*. Birkhäuser, Basel, pp. 213–218. MR: 820223 (cit. on p. 487).

Harry Furstenberg (1967). "Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation". *Math. Systems Theory* 1, pp. 1–49. MR: 0213508 (cit. on pp. 497, 498).

– (1977). "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions". *J. Analyse Math.* 31, pp. 204–256. MR: 0498471.

Hillel Furstenberg and Benjamin Weiss (1996). "A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2} x)$". In: *Convergence in ergodic theory and probability (Columbus, OH, 1993)*. Vol. 5. Ohio State Univ. Math. Res. Inst. Publ. de Gruyter, Berlin, pp. 193–227. MR: 1412607 (cit. on p. 498).

W. T. Gowers (2010). "Decompositions, approximate structure, transference, and the Hahn-Banach theorem". *Bull. Lond. Math. Soc.* 42.4, pp. 573–606. MR: 2669681 (cit. on p. 498).

Ben Green (2014). "Approximate algebraic structure". In: *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. 1*. Kyung Moon Sa, Seoul, pp. 341–367. MR: 3728475 (cit. on p. 487).

Ben Green and Adam J. Harper (2014). "Inverse questions for the large sieve". *Geom. Funct. Anal.* 24.4, pp. 1167–1203. MR: 3248483 (cit. on p. 493).

Larry Guth (2010). "The endpoint case of the Bennett-Carbery-Tao multilinear Kakeya conjecture". *Acta Math.* 205.2, pp. 263–286. MR: 2746348 (cit. on p. 500).

– (2016a). "A restriction estimate using polynomial partitioning". *J. Amer. Math. Soc.* 29.2, pp. 371–413. MR: 3454378 (cit. on p. 500).

– (2016b). "Degree reduction and graininess for Kakeya-type sets in $\mathbb{R}^3$". *Rev. Mat. Iberoam.* 32.2, pp. 447–494. MR: 3512423 (cit. on p. 500).

– (2016c). *Polynomial methods in combinatorics*. Vol. 64. University Lecture Series. American Mathematical Society, Providence, RI, pp. ix+273. MR: 3495952 (cit. on p. 485).

Larry Guth and Nets Hawk Katz (2015). "On the Erdős distinct distances problem in the plane". *Ann. of Math. (2)* 181.1, pp. 155–190. MR: 3272924 (cit. on pp. 488, 490).

Brandon Hanson (2017). "Additive Correlation and the Inverse Problem for the Large Sieve". arXiv: 1706.06958 (cit. on p. 493).

Helmut Hasse (1936). "Zur Theorie der abstrakten elliptischen Funktionenkörper III". *Crelle's Journal* 175, pp. 193–208 (cit. on p. 492).

D. R. Heath-Brown (2002). "The density of rational points on curves and surfaces". *Ann. of Math. (2)* 155.2, pp. 553–595. MR: 1906595 (cit. on p. 494).

H. A. Helfgott and A. Venkatesh (2009). "How small must ill-distributed sets be?" In: *Analytic number theory*. Cambridge Univ. Press, Cambridge, pp. 224–234. MR: 2508647 (cit. on pp. 492, 493).

Harald A. Helfgott (2015). "Growth in groups: ideas and perspectives". *Bull. Amer. Math. Soc. (N.S.)* 52.3, pp. 357–413. MR: 3348442 (cit. on p. 487).

Bernard Host and Bryna Kra (2005). "Nonconventional ergodic averages and nilmanifolds". *Ann. of Math. (2)* 161.1, pp. 397–488. MR: 2150389 (cit. on p. 498).

Henryk Iwaniec and Emmanuel Kowalski (2004). *Analytic number theory*. Vol. 53. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, pp. xii+615. MR: 2061214 (cit. on p. 495).

Aimee S. A. Johnson (1992). "Measures on the circle invariant under multiplication by a nonlacunary subsemigroup of the integers". *Israel J. Math.* 77.1-2, pp. 211–240. MR: 1194793 (cit. on p. 497).

Nets Hawk Katz (2014). "The flecnode polynomial: a central object in incidence geometry". In: *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. III*. Kyung Moon Sa, Seoul, pp. 303–314. MR: 3729029 (cit. on p. 491).

Nets Hawk Katz, Izabella Łaba, and Terence Tao (2000). "An improved bound on the Minkowski dimension of Besicovitch sets in $\mathbf{R}^3$". *Ann. of Math. (2)* 152.2, pp. 383–446. MR: 1804528 (cit. on p. 500).

László Lovász and Balázs Szegedy (2007). "Szemerédi's lemma for the analyst". *Geom. Funct. Anal.* 17.1, pp. 252–270. MR: 2306658 (cit. on p. 498).

J. Milnor (1964). "On the Betti numbers of real varieties". *Proc. Amer. Math. Soc.* 15, pp. 275–280. MR: 0161339 (cit. on p. 499).

Hugh L. Montgomery (1971). *Topics in multiplicative number theory*. Lecture Notes in Mathematics, Vol. 227. Springer-Verlag, Berlin-New York, pp. ix+178. MR: 0337847 (cit. on p. 495).

Yu. V. Nesterenko (1984). "Estimates for the characteristic function of a prime ideal". *Mat. Sb. (N.S.)* 123(165).1, pp. 11–34. MR: 728927 (cit. on p. 498).

János Pach and Micha Sharir (1998). "On the number of incidences between points and curves". *Combin. Probab. Comput.* 7.1, pp. 121–127. MR: 1611057 (cit. on p. 486).

I. G. Petrovskiĭ and O. A. Oleĭnik (1949). "On the topology of real algebraic surfaces". *Izvestiya Akad. Nauk SSSR. Ser. Mat.* 13, pp. 389–402. MR: 0034600 (cit. on p. 490).

Daniel J. Rudolph (1990). "×2 and ×3 invariant measures and entropy". *Ergodic Theory Dynam. Systems* 10.2, pp. 395–406. MR: 1062766 (cit. on p. 497).

Per Salberger (2010). "Counting rational points on projective varieties". preprint (cit. on p. 494).

E. M. Stein (1979). "Some problems in harmonic analysis". In: *Harmonic analysis in Euclidean spaces (Proc. Sympos. Pure Math., Williams Coll., Williamstown, Mass., 1978), Part 1*. Proc. Sympos. Pure Math., XXXV, Part. Amer. Math. Soc., Providence, R.I., pp. 3–20. MR: 545235 (cit. on p. 500).

S. A. Stepanov (1969). "The number of points of a hyperelliptic curve over a finite prime field". *Izv. Akad. Nauk SSSR Ser. Mat.* 33, pp. 1171–1181. MR: 0252400 (cit. on p. 491).

Endre Szemerédi (1978). "Regular partitions of graphs". In: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*. Vol. 260. Colloq. Internat. CNRS. CNRS, Paris, pp. 399–401. MR: 540024 (cit. on p. 498).

Endre Szemerédi and William T. Trotter Jr. (1983). "Extremal problems in discrete geometry". *Combinatorica* 3.3-4, pp. 381–392. MR: 729791 (cit. on p. 486).

Terence Tao (2014). "Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory". *EMS Surv. Math. Sci.* 1.1, pp. 1–46. MR: 3200226 (cit. on p. 485).

René Thom (1965). "Sur l'homologie des variétés algébriques réelles". In: *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*. Princeton Univ. Press, Princeton, N.J., pp. 255–265. MR: 0200942 (cit. on p. 499).

M. Walsh (n.d.). *Polynomial partitioning over varieties*. Preprint (cit. on p. 499).

Miguel N. Walsh (2012a). "Norm convergence of nilpotent ergodic averages". *Ann. of Math. (2)* 175.3, pp. 1667–1688. MR: 2912715 (cit. on p. 498).

– (2012b). "The inverse sieve problem in high dimensions". *Duke Math. J.* 161.10, pp. 2001–2022. MR: 2954623 (cit. on pp. 489, 493, 498).

– (2014). "The algebraicity of ill-distributed sets". *Geom. Funct. Anal.* 24.3, pp. 959–967. MR: 3213836 (cit. on pp. 493, 498).

– (2015). "Bounded rational points on curves". *Int. Math. Res. Not. IMRN* 14, pp. 5644–5658. MR: 3384452 (cit. on p. 495).

André Weil (1949). "Numbers of solutions of equations in finite fields". *Bull. Amer. Math. Soc.* 55, pp. 497–508. MR: 0029393 (cit. on p. 492).

Thomas Wolff (1999). "Recent work connected with the Kakeya problem". In: *Prospects in mathematics (Princeton, NJ, 1996)*. Amer. Math. Soc., Providence, RI, pp. 129–162. MR: 1660476 (cit. on p. 488).

Trevor D. Wooley (2012). "Vinogradov's mean value theorem via efficient congruencing". *Ann. of Math. (2)* 175.3, pp. 1575–1627. MR: 2912712 (cit. on p. 496).

Received 2017-12-03.

Miguel N. Walsh
Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
1428 Buenos Aires
Argentina
walsh@maths.ox.ac.uk
mwalsh@dm.uba.ar

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (505–540)

PERIODS, CYCLES, AND L-FUNCTIONS: A RELATIVE TRACE FORMULA APPROACH

Wei Zhang (张伟)

Abstract

This is a report for the author's talk in ICM-2018. Motivated by the formulas of Gross–Zagier and Waldspurger, we review conjectures and theorems on automorphic period integrals, special cycles on Shimura varieties, and their connection to central values of L-functions and their derivatives. We focus on the global Gan–Gross–Prasad conjectures, their arithmetic versions and some variants in the author's joint work with Rapoport and Smithling. We discuss the approach of relative trace formulas and the arithmetic fundamental lemma conjecture. In the function field setting, Z. Yun and the author obtain a formula for higher order derivatives of L-functions in terms of special cycles on the moduli space of Drinfeld Shtukas.

Contents

Research of W. Zhang partially supported by the NSF grant DMS #1601144, and a Simmons fellowship.
MSC2010: primary 11F67; secondary 11F70, 11G40, 14C25, 14G35, 14H60.
Keywords: L-functions, Shimura varieties, Drinfeld Shtukas, Special cycles, Relative trace formula, Gross–Zagier formula, Waldspurger formula, Rapoport–Zink spaces, Spherical subgroup.

1 Introduction

We begin with a special case of Pell's equation

$$x^2 - py^2 = 1, \quad x, y \in \mathbb{Z},$$

where $p \equiv 1 \bmod 4$ is a prime number. Let K be the real quadratic field $\mathbb{Q}[\sqrt{p}]$ and O_K its ring of integers. Then most solutions to Pell's equation can be extracted from the group of units in O_K, which is known to be the product of $\{\pm 1\}$ and an infinite cyclic group generated by a fundamental unit. In 1830s, Dirichlet systematically constructed units in O_K using special values of trigonometric functions:

$$\theta_p = \frac{\prod_{a \not\equiv \square \bmod p} \sin \frac{a\pi}{p}}{\prod_{b \equiv \square \bmod p} \sin \frac{b\pi}{p}}, \quad 0 < a, b < p/2,$$

where $\square \bmod p$ denotes a square residue. Dirichlet also showed that the obstruction for θ_p to be a fundamental unit is the class group of K, with the help of an infinite series

$$L\left(s, \left(\frac{\cdot}{p}\right)\right) = \sum_{n \geq 1,\, p \nmid n} \left(\frac{n}{p}\right) n^{-s}, \quad s \in \mathbb{C}, \quad \mathrm{Re}(s) > 1.$$

Here $\left(\frac{\cdot}{p}\right)$ denotes the Legendre symbol for quadratic residues. This is now called a Dirichlet L-series, and it has holomorphic continuation to $s \in \mathbb{C}$ with a simple zero at $s = 0$. What Dirichlet discovered can be stated as two formulas for the value at $s = 0$ of the first derivative: the first one is in terms of θ_p,

$$L'\left(0, \left(\frac{\cdot}{p}\right)\right) = \log \theta_p, \tag{1-1}$$

and the second one is in terms of the class number h_p and the fundamental unit $\epsilon_p > 1$,

$$L'\left(0, \left(\frac{\cdot}{p}\right)\right) = h_p \log \epsilon_p. \tag{1-2}$$

Dirichlet also proved two formulas for an imaginary quadratic field. For simplicity, let $p \equiv 7 \bmod 8$ be a prime. Now the L-series $L\left(s, \left(\frac{\cdot}{p}\right)\right)$ does not vanish at $s = 0$. His first formula states

$$L\left(0, \left(\frac{\cdot}{p}\right)\right) = \sum_{0 < a < p/2} \left(\frac{a}{p}\right), \tag{1-3}$$

and the second one is in terms of the class number h_{-p} of $\mathbb{Q}[\sqrt{-p}]$

$$L\left(0,\left(\frac{\cdot}{p}\right)\right)=h_{-p}. \tag{1-4}$$

A non-trivial corollary is a finite expression for the class number

$$h_{-p}=\sum_{0<a<p/2}\left(\frac{a}{p}\right),$$

i.e., the difference between the number of square residues and of non-square residues in the interval $(0,p/2)$.

In 1952, Heegner discovered a way to construct rational points (over some number fields) on elliptic curves using special values of modular functions, in a manner similar to Dirichlet's solutions to Pell's equation. For instance, the elliptic curve

$$E:y^2=x^3-1728$$

is parameterized by modular functions (γ_2,γ_3), where

$$\gamma_2(z)=\frac{E_4}{\eta^8}=\frac{1+240\sum_{n=1}^{\infty}\sigma_3(n)q^n}{q^{1/3}\prod_{n=1}^{\infty}(1-q^n)^8},\qquad \gamma_3(z)=\frac{E_6}{\eta^{12}}=\frac{1-504\sum_{n=1}^{\infty}\sigma_5(n)q^n}{q^{1/2}\prod_{n=1}^{\infty}(1-q^n)^{12}}.$$

Here as customary, $z\in\mathcal{H}$ is on the upper half plane, and $q=e^{2\pi iz}$. Then Heegner's strategy is to evaluate (γ_2,γ_3) at z in an auxiliary imaginary quadratic field K to construct points of E with coordinates in an abelian extension of K. The theorem of Gross and Zagier [1986] then relates the Néron–Tate heights of Heegner's points to special values of the *first order* derivatives of certain L-functions, providing an analog of Dirichlet's formula for a real quadratic field (1-1). In this new context, an analog of Dirichlet's formula for an imaginary quadratic field (1-3) is the theorem of Waldspurger [1985] relating toric period integrals (cf. Section 2.2.1) to special values of L-functions of the same sort as in the work of Gross and Zagier.

A natural question is to generalize the constructions of Dirichlet and of Heegner to higher dimensional algebraic varieties, and at the same time to generalize their relation to appropriate L-values. As a partial answer to this question, in this report we will consider some special algebraic cycles on Shimura varieties, cf. Section 3.1.1. A class of such special cycles are the arithmetic diagonal cycles that appear in the arithmetic Gan–Gross–Prasad conjecture Gan, Gross, and Prasad [2012, §27].

A parallel question is the generalization of the formulas of Dirichlet (1-3) and of Waldspurger to automorphic L-functions on a higher rank reductive group G over a global field

F. In this direction, we will consider automorphic period integrals for a spherical subgroup H of G, i.e.,

$$\int_{\mathrm{H}(F)\backslash \mathrm{H}(\mathbb{A})} \phi(h)\, dh,$$

where ϕ is a function on $\mathrm{G}(F)\backslash\mathrm{G}(\mathbb{A})$, and $\mathbb{A}$ the the ring of adèles of F. The central object is the quotient $\mathrm{H}(F)\backslash\mathrm{H}(\mathbb{A})$ sitting inside $\mathrm{G}(F)\backslash\mathrm{G}(\mathbb{A})$. This paradigm may be viewed as a degenerate case of special cycles sitting inside the ambient Shimura variety.

To study the relationship between automorphic periods and L-values, Jacquet invented the relative trace formula, cf. Jacquet [2005]. In W. Zhang [2012b] we adopted the relative trace formula approach to study height pairings of special cycles in the arithmetic Gan–Gross–Prasad conjecture. In this context, we formulated local conjectures (on intersection numbers of the arithmetic diagonal cycle on Rapoport–Zink spaces), namely the *arithmetic fundamental lemma* conjecture (cf. W. Zhang [ibid.]), and the *arithmetic transfer* conjecture by Rapoport, Smithling and the author (cf. Rapoport, Smithling, and W. Zhang [2017a] and Rapoport, Smithling, and W. Zhang [2018]). In this report we will review the approach, the conjectures and the status.

Another natural question is in the direction of higher order derivatives of L-functions in the Gross–Zagier formula. In Yun and W. Zhang [2017] Yun and the author found a geometric interpretation of the higher order derivative in the functional field setting, in terms of special cycles on the moduli stack of Drinfeld Shtukas with multiple number of modifications, cf. Section 4. To generalize this to the number field case, one is led to the tantalizing question of finding Shtukas over number fields.

Limited by the length of this report, we will not discuss the analog of the class number formula (1-2) and (1-4) in our setting. In the context of elliptic curves over $\mathbb{Q}$, this is the conjectural formula of Birch and Swinnerton–Dyer. When the analytic rank is one (resp. zero), we have an equivalent statement in terms of the divisibility of Heegner points (resp. of normalized toric period integrals), thanks to the formula of Gross–Zagier (resp. of Waldspurger). Much of the equivalent statement has been proved in the past thirty years. Beyond elliptic curves (or modular forms on GL_2), there have been many recent developments where special cycles play a crucial role in the study of Selmer groups of Bloch–Kato type. We hope to return to this topic on another occasion.

Acknowledgments. The author thanks B. Gross, M. Rapoport, Y. Sakellaridis, B. Smithling and S. Zhang for helpful comments.

Notation. Let F be a global field (unless otherwise stated), i.e., a number filed or a function field (of a geometrically connected smooth proper curve X over a finite field $k = \mathbb{F}_q$). Let $\mathbb{A} = \mathbb{A}_F = \prod_v' F_v$ be the ring of adèles, the restricted direct product over all completions F_v of F. The ring of integers in a non-archimedean local field F_v is denoted by O_{F_v}. For a subset of places S, we let $\mathbb{A}_S = \prod'_{v\in S} F_v$. When F is a number field, we write $\mathbb{A} = \mathbb{A}_f \times \mathbb{A}_\infty$, where $\mathbb{A}_f$ is the ring of the finite adèles and $\mathbb{A}_\infty = F \otimes_{\mathbb{Q}} \mathbb{R}$.

For a field extension F'/F and an algebraic group G over F', we denote by $\mathrm{R}_{F'/F}\mathrm{G}$ the Weil restriction of scalars. We denote by $\mathbb{G}_m$ the multiplicative group.

2 Automorphic periods and L-values

2.1 Spherical pairs and automorphic periods.

2.1.1 Automorphic quotient. Let G be an algebraic group over F. We define the *automorphic quotient* associated to G to be the quotient topological space

$$[\mathrm{G}] := \mathrm{G}(F)\backslash\mathrm{G}(\mathbb{A}). \tag{2-1}$$

Let $K \subset \mathrm{G}(\mathbb{A})$ be a subgroup. When F is a function field, we assume that K is a compact open subgroup of $\mathrm{G}(\mathbb{A})$. When F is a number field, we assume that K is a product $K_f \cdot K_\infty$ where $K_f \subset \mathrm{G}(\mathbb{A}_f)$ is a compact open subgroup, and K_∞ is a suitable subgroup of $\mathrm{G}(\mathbb{A}_\infty)$. We then define a quotient

$$[\mathrm{G}]_K := \mathrm{G}(F)\backslash\left[\mathrm{G}(\mathbb{A})/K\right]. \tag{2-2}$$

When F is a function field, this is a discrete set (or naturally as a groupoid). When $F = k(X)$, $\mathrm{G} = \mathrm{GL}_n$, and $K = \prod_v \mathrm{GL}_n(O_{F_v})$, the groupoid $[\mathrm{G}]_K$ is naturally isomorphic to the groupoid $\mathrm{Bun}_n(k)$, the k-points of Bun_n (the stack of vector bundles of rank n on X).

2.1.2 Automorphic period. From now on let G be a reductive group over F. Let $\mathrm{H} \subset \mathrm{G}$ be a subgroup. Let Z_{G} be the center of G and let $Z = \mathrm{H} \cap Z_{\mathrm{G}}$. Let $\mathcal{Q}_0(\mathrm{G})$ be the space of cuspidal automorphic forms on $[\mathrm{G}]$, invariant under the action of $Z(\mathbb{A})$. Then the *automorphic H-period integral* is defined by

$$\begin{aligned} \mathcal{P}_{\mathrm{H}}\colon \mathcal{Q}_0(\mathrm{G}) &\longrightarrow \mathbb{C} \\ \phi &\longmapsto \int_{Z(\mathbb{A})\backslash[\mathrm{H}]} \phi(h)\, dh. \end{aligned}$$

Remark 2.1. The name "automorphic period" is different from the period in the context of comparison theorems between various cohomology theories. However, some special cases of the automorphic period integrals may yield periods in the de Rham–Betti comparison theorem. For instance, this happens when the integral can be turned into the form

$\int_{Z(\mathbb{A})\backslash[\mathrm{H}]_{K_{\mathrm{H}}}} \omega_\phi$ for a closed differential form ω_ϕ on the real manifold $[\mathrm{G}]_K$ for suitable K, and $K_{\mathrm{H}} = K \cap \mathrm{H}(\mathbb{A})$.

Let π be a (unitary throughout the article) cuspidal automorphic representation of $\mathrm{G}(\mathbb{A})$. For simplicity, we assume that there is a unique embedding $\pi \hookrightarrow \mathcal{A}_0(\mathrm{G})$ (in many applications this is the case). We consider the restriction of $\mathcal{P}_{\mathrm{H}}$ to π; this defines an element in $\mathcal{P}_{\mathrm{H},\pi} \in \mathrm{Hom}_{\mathrm{H}(\mathbb{A})}(\pi, \mathbb{C})$.

Definition 2.2 (Jacquet)**.** A cuspidal automorphic representation π is *(globally) distinguished by* H if the linear functional $\mathcal{P}_{\mathrm{H},\pi} \in \mathrm{Hom}_{\mathrm{H}(\mathbb{A})}(\pi, \mathbb{C})$ does not vanish, i.e., there exists some $\phi \in \pi$ such that $\mathcal{P}_{\mathrm{H}}(\phi) \neq 0$.

It is also natural to consider a twisted version. Let χ be a character of $Z(\mathbb{A})\mathrm{H}(F)\backslash\mathrm{H}(\mathbb{A})$. We define the automorphic (H, χ)-period integral $\mathcal{P}_{\mathrm{H},\chi}$ in a similar manner,

$$\mathcal{P}_{\mathrm{H},\chi}(\phi) = \int_{Z(\mathbb{A})\backslash[\mathrm{H}]} \phi(h)\chi(h)\, dh.$$

If π is distinguished by H, then $\mathrm{Hom}_{\mathrm{H}(\mathbb{A})}(\pi, \mathbb{C}) \neq 0$, and in particular, $\mathrm{Hom}_{\mathrm{H}(F_v)}(\pi_v, \mathbb{C}) \neq 0$ for every place v. We then say that π_v is (locally) distinguished by $\mathrm{H}(F_v)$ if $\mathrm{Hom}_{\mathrm{H}(F_v)}(\pi_v, \mathbb{C}) \neq 0$. Very often the automorphic period integral $\mathcal{P}_{\mathrm{H}}$ behaves nicely only when the pair (H, G) satisfies certain nice properties, such as

(i) the multiplicity-one property $\dim \mathrm{Hom}_{\mathrm{H}(F_v)}(\pi_v, \mathbb{C}) \leq 1$ holds for all v, or the multiplicity can be described in a certain nice way, and

(ii) the locally distinguished representations can be characterized in terms of L-parameters.

Jacquet and his school have studied the distinction for many instances (locally and globally, cf. Jacquet [2005]). A large class of (H, G) called *spherical pairs* are expected to have the above nice properties, according to the work of Sakellaridis [2008] and his joint work with Sakellaridis and Venkatesh [2012].

2.1.3 Spherical pairs. Let (H, G) be over a field F (now arbitrary). The reductive group G acts on G/H by left multiplication. If F is an algebraically closed field, we say that the pair (H, G) is *spherical* if a Borel subgroup B of G has an open dense orbit on G/H Sakellaridis [2008]. Over a general field F, the pair (H, G) is said to be *spherical* if its base change to an algebraic closure of F is spherical. We then call H a spherical subgroup of G.

Here are some examples.

(i) Whittaker pair (N, G), where G is quasi-split, and N is a maximal nilpotent subgroup.

(ii) The pair $(\mathrm{G}, \mathrm{G} \times \mathrm{G})$ with the diagonal embedding.

(iii) Symmetric pair (H, G), where H is the fixed point locus of an involution $\theta : \mathrm{G} \to \mathrm{G}$. This constitutes a large class of spherical pairs, including

- the unitary periods of Jacquet and Ye (cf. the survey Offen [2009]), where $\mathrm{G} = \mathrm{R}_{F'/F}\mathrm{GL}_n$, and H is a unitary group attached to a Hermitian space with respect to a quadratic extension F'/F. This period is related to the quadratic base change for the general linear group.
- the Flicker–Rallis periods W. Zhang [2014a, §3.2], where $\mathrm{G} = \mathrm{R}_{F'/F}\mathrm{GL}_n$, and $\mathrm{H} = \mathrm{GL}_n$, for quadratic F'/F. This period is related to the quadratic base change for the unitary group.
- the linear periods of Friedberg–Jacquet Friedberg and Jacquet [1993], where $\mathrm{H} = \mathrm{GL}_{n/2} \times \mathrm{GL}_{n/2}$ embeds in $\mathrm{G} = \mathrm{GL}_n$ by

$$(a, d) \longmapsto \mathrm{diag}(a, d).$$

(iv) the Rankin–Selberg pair (named after its connection to Rankin–Selberg convolution L-functions), where $\mathrm{G} = \mathrm{GL}_{n-1} \times \mathrm{GL}_n$, and $\mathrm{H} = \mathrm{GL}_{n-1} \hookrightarrow G$ with the embedding

$$g \longmapsto (g,\ \mathrm{diag}(g, 1)).$$

(There are also spherical subgroups of $\mathrm{GL}_m \times \mathrm{GL}_n$ when $|n - m| > 1$ involving non-reductive subgroups.)

(v) the Gan–Gross–Prasad pairs $(\mathrm{SO}_{n-1}, \mathrm{SO}_{n-1} \times \mathrm{SO}_n)$ and $(\mathrm{U}_{n-1}, \mathrm{U}_{n-1} \times \mathrm{U}_n)$; see Section 2.2.2. They resemble the Rankin–Selberg pairs. (There are also spherical subgroups of $\mathrm{SO}_m \times \mathrm{SO}_n$ and $\mathrm{U}_m \times \mathrm{U}_n$ when $|n - m| > 1$ involving non-reductive subgroups).

2.2 The global Gan–Gross–Prasad conjecture and the Ichino–Ikeda refinement. Let F be a number field for the rest of this section.

2.2.1 Waldspurger formula. Let B be a quaternion algebra over F and let $\mathrm{G} = B^\times$ (as an F-algebraic group). Let F'/F be a quadratic extension of number fields and denote by T the torus $\mathrm{R}_{F'/F}\mathbb{G}_m$. Let $F' \hookrightarrow B$ be an embedding of F-algebras, and $\mathrm{T} \hookrightarrow \mathrm{G}$ the induced embedding of F-algebraic groups. Then (T, G) is a spherical pair. In Waldspurger [1985], Waldspurger studied the automorphic period integral $\mathcal{P}_{T,\chi}$ (sometimes called the *toric* period), and he proved an exact formula relating the square $|\mathcal{P}_{T,\chi}|^2$ to a certain central L-value.

Below we will consider one generalization of Waldspurger's formula to higher rank groups, i.e., the global conjectures of Gan–Gross–Prasad and Ichino–Ikeda. For this report, we implicitly assume the endoscopic classification of Arthur for orthogonal and unitary groups.

2.2.2 The global Gan–Gross–Prasad conjecture. Gan, Gross, and Prasad [2012] proposed a series of precise conjectures regarding the local and global distinction for (H, G) when G is a classical group (orthogonal, unitary and symplectic), extending the conjectures of Gross and Prasad [1992, 1994] for orthogonal groups.

We recall their global conjectures in orthogonal and Hermitian cases. For simplicity, we restrict to the case when the spherical subgroup is reductive. Let F be a number field, and let $F' = F$ in the orthogonal case and F' a quadratic extension of F in the Hermitian case. Let W_n be a non-degenerate orthogonal space or Hermitian space with F'-dimension n. Let $W_{n-1} \subset W_n$ be a non-degenerate subspace of codimension one. Let G_i be $\mathrm{SO}(W_i)$ or $\mathrm{U}(W_i)$ for $i = n-1, n$, and $\delta : \mathrm{G}_{n-1} \hookrightarrow \mathrm{G}_n$ the induced embedding. Let

$$\mathrm{G} = \mathrm{G}_{n-1} \times \mathrm{G}_n, \qquad \mathrm{H} = \mathrm{G}_{n-1}, \tag{2-3}$$

with the "diagonal" embedding $\Delta : \mathrm{H} \hookrightarrow \mathrm{G}$ (i.e., the graph of δ). The pair (H, G) is spherical and we call it the *Gan–Gross–Prasad pair*.

Let $\pi = \pi_{n-1} \boxtimes \pi_n$ be a tempered cuspidal automorphic representation of $\mathrm{G}(\mathbb{A})$. The central L-values of certain automorphic L-functions $L(s, \pi, R)$ show up in their conjecture, where R is a finite dimensional representation of the L-group ${}^L\mathrm{G}$, cf. Gan, Gross, and Prasad [2012, §22]. We can describe the L-function as the Rankin–Selberg convolution of suitable automorphic representations on general linear groups. For $i \in \{n-1, n\}$, let $\Pi_{i,F'}$ be the endoscopic functoriality transfer of π_i from G_i to suitable $\mathrm{GL}_N(\mathbb{A}_{F'})$: in the Hermitian case, this is the base change of π_i to $\mathrm{GL}_i(\mathbb{A}_{F'})$; and in the orthogonal case, this is the endoscopic transfer from $\mathrm{G}_i(\mathbb{A})$ to $\mathrm{GL}_i(\mathbb{A})$ (resp. $\mathrm{GL}_{i-1}(\mathbb{A})$) if i is even (resp. odd). Then the L-function $L(s, \pi, R)$ can be defined more explicitly as the Rankin–Selberg convolution L-function $L(s, \Pi_{n-1,F'} \times \Pi_{n,F'})$.

We are ready to state the global Gan–Gross–Prasad conjecture Gan, Gross, and Prasad [ibid., §24].

Conjecture 2.3. *Let π be a tempered cuspidal automorphic representation of $\mathrm{G}(\mathbb{A})$. The following statements are equivalent.*

(i) *The automorphic* H*-period integral does not vanish on π, i.e., $\mathcal{P}_{\mathrm{H}}(\phi) \neq 0$ for some $\phi \in \pi$.*

(ii) *The space $\mathrm{Hom}_{\mathrm{H}(\mathbb{A})}(\pi, \mathbb{C}) \neq 0$ and the central value $L(\frac{1}{2}, \pi, R) \neq 0$.*

Remark 2.4. It is known that the pair $(\mathrm{H}(F_v), \mathrm{G}(F_v))$ satisfies the multiplicity one property by Aizenbud, Gourevitch, Rallis, and Schiffmann [2010] for p-adic local fields, and by Sun and C.-B. Zhu [2012] for archimedean local fields. The local conjectures of Gan, Gross, and Prasad Gan, Gross, and Prasad [2012, §17] specify the member π_v in a generic Vogan L-packet (cf. Gan, Gross, and Prasad [ibid., §9-11]) with $\dim \mathrm{Hom}_{\mathrm{H}(F_v)}(\pi_v, \mathbb{C}) =$

1, in terms of local root numbers associated to the L-parameter. Their local conjectures are mostly proved by Gan, Gross, Prasad, and Waldspurger [2012] and Mœglin and Waldspurger [2012] (p-adic orthogonal groups), and Beuzart-Plessis [2015a] and Beuzart-Plessis [2015b] (unitary groups over p-adic and archimedean local fields). He [2017] gives an alternative proof for discrete series representations of unitary groups over archimedean local fields.

We have the following result.

Theorem 2.5. *Let* $\mathrm{G} = \mathrm{U}(W_{n-1}) \times \mathrm{U}(W_n)$ *for Hermitian spaces* $W_{n-1} \subset W_n$ *over a quadratic extension* F' *of* F. *Let* π *be a tempered cuspidal automorphic representation of* $\mathrm{G}(\mathbb{A})$ *such that there exist a non-archimedean place* v *of* F *split in* F' *where* π_v *is supercuspidal. Then Conjecture 2.3 holds.*

This was proved in W. Zhang [2014b] under a further condition on the archimedean places, which was later removed by Xue [n.d.]. The local condition above was due to a simple version of the Jacquet–Rallis relative trace formula in W. Zhang [2014b], cf. Section 5.2. Chaudouard and Zydor [2016] and Zydor [2015] have made progress towards the full relative trace formula that should remove the local condition.

Remark 2.6. Ginzburg, Jiang, and Rallis [2004, 2009] have proved the direction (i) $\Longrightarrow$ (ii) of Conjecture 2.3 for both the orthogonal and Hermitian cases when the group G is quasi-split and the representation π is (globally) generic (cf. Jiang and L. Zhang [2015, Theorem 5.7].

2.2.3 Ichino–Ikeda refinement. For many applications, we would like to have a refined version of the Gan–Gross–Prasad conjecture, analogous to the Waldspurger formula for the toric period $\mathcal{P}_{\mathrm{T},\chi}$ in Section 2.2.1. We recall the refinement of Ichino and Ikeda [2010] (for orthogonal groups; later their idea was carried out for unitary groups by N. Harris in Harris [2014]).

Let $L(s, \pi, \mathrm{Ad})$ be the adjoint L-function (cf. Gan, Gross, and Prasad [2012, §7]). Denote $\Delta_n = L(M^\vee(1))$ where $M^\vee$ is the motive dual to the motive M associated to G_n by Gross [1997]. It is a product of special values of Artin L-functions. We will be interested in the following combination of L-functions,

$$\mathcal{L}(s, \pi) = \Delta_n \frac{L(s, \pi, R)}{L(1, \pi, \mathrm{Ad})}. \tag{2-4}$$

We also write $\mathcal{L}(s, \pi_v)$ for the corresponding local factor at a place v.

Let π_v be an irreducible tempered unitary representation of $\mathrm{G}(F_v)$ with an invariant inner product $\langle \cdot, \cdot \rangle_v$. Ichino and Ikeda construct a canonical element in the space

$\mathrm{Hom}_{\mathrm{H}(F_v)}(\pi_v, \mathbb{C}) \otimes \mathrm{Hom}_{\mathrm{H}(F_v)}(\overline{\pi}_v, \mathbb{C})$ by integrating matrix coefficients: for $\phi_v, \varphi_v \in \pi_v$,

$$\widetilde{\alpha}_v(\phi_v, \varphi_v) = \int_{\mathrm{H}(F_v)} \langle \pi_v(h)\phi_v, \varphi_v\rangle_v \, dh. \tag{2-5}$$

Ichino and Ikeda showed that the integral converges absolutely for all (tempered) π. In fact, the convergence holds for any *strongly tempered* pair (H, G), cf. Sakellaridis and Venkatesh [2012]. When π_v is unramified and the vectors ϕ_v, φ_v are fixed by a hyperspecial compact open $\mathrm{G}(O_{F_v})$ such that $\langle \phi_v, \varphi_v\rangle_v = 1$, we have

$$\widetilde{\alpha}_v(\phi_v, \varphi_v) = \mathscr{L}(\frac{1}{2}, \pi_v) \cdot \mathrm{vol}(\mathrm{H}(O_{F_v})).$$

We normalize the local canonical invariant form:

$$\alpha_v(\phi_v, \varphi_v) = \frac{1}{\mathscr{L}(\frac{1}{2}, \pi_v)} \widetilde{\alpha}_v(\phi_v, \varphi_v). \tag{2-6}$$

We endow $\mathrm{H}(\mathbb{A})$ (resp. $\mathrm{G}(\mathbb{A})$) with their Tamagawa measures and $[\mathrm{H}]$ (resp. $[\mathrm{G}]$) with the quotient measure by the counting measure on $\mathrm{H}(F)$ (resp. $\mathrm{G}(F)$). We choose the Haar measure dh on $\mathrm{H}(\mathbb{A})$ and the measures dh_v on $\mathrm{H}(F_v)$ such that $dh = \prod_v dh_v$. Let $\langle \phi, \varphi\rangle_{\mathrm{Pet}}$ be the Petersson inner product of $\phi, \varphi \in \pi = \otimes_v \pi_v$, and choose the local inner products $\langle \cdot, \cdot\rangle_v$ such that $\langle \phi, \varphi\rangle = \prod_v \langle \phi_v, \varphi_v\rangle_v$ for $\phi = \otimes_v \phi_v$ and $\varphi = \otimes_v \varphi_v$.

We can now state the Ichino–Ikeda conjecture Ichino and Ikeda [2010, Conj. 1.5 and 2.1] that refines the global Gan–Gross–Prasad Conjecture 2.3.

Conjecture 2.7. *Let π be a tempered cuspidal automorphic representation of* $\mathrm{G}(\mathbb{A})$. *Then for $\phi = \otimes_v \phi_v \in \pi$,*

$$\left|\mathscr{P}_{\mathrm{H}}(\phi)\right|^2 = 2^{-\beta_\pi} \mathscr{L}(1/2, \pi) \prod_v \alpha_v(\phi_v, \phi_v), \tag{2-7}$$

where β_π is the rank of a finite elementary 2*-group associated to the L-parameter of π.*

Remark 2.8. If $\mathrm{Hom}_{\mathrm{H}(\mathbb{A})}(\pi, \mathbb{C}) = 0$, both sides of (2-7) vanish.

Remark 2.9. In the orthogonal case and $n = 3$, the refined conjecture is exactly the same as the formula of Waldspurger. When $n = 4$ the conjecture is proved by Ichino [2008]. Little is known in the higher rank case, cf. the survey Gan [2014].

In the Hermitian case and $n = 2$, the conjecture follows from Waldspurger's formula Harris [2014]. In general, we have the following result, due to the author W. Zhang [2014a] and Beuzart-Plessis [2016].

Theorem 2.10. *Let* $\mathrm{G} = \mathrm{U}(W_{n-1}) \times \mathrm{U}(W_n)$ *for Hermitian spaces* $W_{n-1} \subset W_n$ *over a quadratic extension* F' *of* F. *Let* π *be a tempered cuspidal automorphic representation of* $\mathrm{G}(\mathbb{A})$. *Assume that*

(i) *there exists a non-archimedean place* v *of* F *split in* F' *such that* π_v *is supercuspidal, and*

(ii) *all archimedean places of* F *are split in* F'.

Then Conjecture 2.7 holds.

For some ingredients of the proof, see Section 5.2.3.

2.2.4 Reformulation in terms of spherical characters. Let π be a cuspidal automorphic representation of $\mathrm{G}(\mathbb{A})$. We define the *global spherical character* $\mathbb{I}_\pi$ as the distribution on $\mathrm{G}(\mathbb{A})$

$$\mathbb{I}_\pi(f) := \sum_{\phi \in \mathrm{OB}(\pi)} \mathscr{P}_{\mathrm{H}}(\pi(f)\phi)\overline{\mathscr{P}_{\mathrm{H}}(\phi)}, \quad f \in \mathscr{C}_c^\infty(\mathrm{G}(\mathbb{A})), \tag{2-8}$$

where the sum runs over an orthonormal basis $\mathrm{OB}(\pi)$ of π (with respect to the Petersson inner product). Note that $\mathbb{I}_\pi$ is an *eigen-distribution* for the spherical Hecke algebra $\mathscr{H}^S(\mathrm{G})$ away from a sufficiently large set S (including all bad primes), in the sense that for all $f = f_S \otimes f^S$ with $f^S \in \mathscr{H}^S(\mathrm{G})$ and $f_S \in \mathscr{C}_c^\infty(\mathrm{G}(\mathbb{A}_S))$,

$$\mathbb{I}_\pi(f) = \lambda_{\pi^S}\left(f^S\right) \mathbb{I}_\pi\left(f_S \otimes 1_{K^S}\right), \tag{2-9}$$

where λ_{π^S} is the "eigen-character" of $\mathscr{H}^S(\mathrm{G})$ associated to π^S.

We define the *local spherical character* in terms of the local canonical invariant form α_v in (2-6),

$$\mathbb{I}_{\pi_v}(f_v) := \sum_{\phi_v \in \mathrm{OB}(\pi_v)} \alpha_v(\pi_v(f_v)\phi_v, \phi_v), \quad f_v \in \mathscr{C}_c^\infty(\mathrm{G}(F_v)), \tag{2-10}$$

where the sum runs over an orthonormal basis $\mathrm{OB}(\pi_v)$ of π_v.

In W. Zhang [2014a, Conj. 1.6] the author stated an alternative version of the Ichino–Ikeda conjecture in terms of spherical characters.

Conjecture 2.11. *Let* π *be a tempered cuspidal automorphic representation of* $\mathrm{G}(\mathbb{A})$. *Then for all pure tensors* $f = \bigotimes_v f_v \in \mathscr{C}_c^\infty(\mathrm{G}(\mathbb{A}))$,

$$\mathbb{I}_\pi(f) = 2^{-\beta_\pi} \mathscr{L}(1/2, \pi) \prod_v \mathbb{I}_{\pi_v}(f_v).$$

By W. Zhang [2014a, Lemma 1.7], this conjecture is equivalent to Conjecture 2.7. This new formulation is more suitable for the relative trace formula approach, cf. Section 5.2.3. This also inspires us to state a version of the refined arithmetic Gan–Gross–Prasad conjecture where the Ichino–Ikeda formulation does not seem to apply directly, cf. Section 3.2.3, Conjecture 3.5.

3 Special cycles and L-derivatives

3.1 Special pairs of Shimura data and special cycles.

3.1.1 . To describe our set-up, we introduce the concept of *a special pair of Shimura data*. Let $\mathbb{S}$ be the torus $\mathrm{R}_{\mathbb{C}/\mathbb{R}}\mathbb{G}_m$ over $\mathbb{R}$ (i.e., we view $\mathbb{C}^\times$ as an $\mathbb{R}$-group). Recall that a Shimura datum $\big(\mathrm{G}, X_{\mathrm{G}}\big)$ consists of a reductive group G over $\mathbb{Q}$, and a $\mathrm{G}(\mathbb{R})$-conjugacy class $X_{\mathrm{G}} = \{h_{\mathrm{G}}\}$ of $\mathbb{R}$-group homomorphisms $h_{\mathrm{G}} : \mathbb{S} \to \mathrm{G}_{\mathbb{R}}$ (sometimes called Shimura homomorphisms) satisfying Deligne's list of axioms Deligne [1971, p. 1.5]. In particular, X_{G} is a Hermitian symmetric domain.

Definition 3.1. A *special pair* of Shimura data is a homomorphism Deligne [ibid., p. 1.14] between two Shimura data

$$\delta\colon \big(\mathrm{H}, X_{\mathrm{H}}\big) \longrightarrow \big(\mathrm{G}, X_{\mathrm{G}}\big)$$

such that

(i) the homomorphism $\delta : \mathrm{H} \to \mathrm{G}$ is injective such that the pair (H, G) is spherical, and

(ii) the dimensions of X_{H} and X_{G} (as complex manifolds) satisfy

$$\dim_{\mathbb{C}} X_{\mathrm{H}} = \left\lfloor \frac{\dim_{\mathbb{C}} X_{\mathrm{G}}}{2} \right\rfloor.$$

In particular, we enhance a spherical pair (H, G) to a homomorphism of Shimura data $(\mathrm{H}, X_{\mathrm{H}}) \to (\mathrm{G}, X_{\mathrm{G}})$.

Remark 3.2. It seems an interesting question to enumerate special pairs of Shimura data. In fact, we may consider the analog of special pairs of Shimura data in the context of *local Shimura data* Rapoport and Viehmann [2014]. It seems more realistic to enumerate the pairs in the local situation.

For a Shimura datum $(\mathrm{G}, X_{\mathrm{G}})$ we have a projective system of *Shimura varieties* $\{\mathrm{Sh}_K(\mathrm{G})\}$, indexed by compact open subgroups $K \subset \mathrm{G}(\mathbb{A}_f)$, of smooth quasi-projective varieties (for neat K) defined over a number field E—the reflex field of $(\mathrm{G}, X_{\mathrm{G}})$.

For a special pair of Shimura data $(\mathrm{H}, X_{\mathrm{H}}) \to (\mathrm{G}, X_{\mathrm{G}})$, compact open subgroups $K_{\mathrm{H}} \subset \mathrm{H}(\mathbb{A}_f)$ and $K_{\mathrm{G}} \subset \mathrm{G}(\mathbb{A}_f)$ such that $K_{\mathrm{H}} \subset K_{\mathrm{G}}$, we have a finite morphism (over the reflex field of $(\mathrm{H}, X_{\mathrm{H}})$)

$$\delta_{K_{\mathrm{H}},K_{\mathrm{G}}} : \mathrm{Sh}_{K_{\mathrm{H}}}(\mathrm{H}) \longrightarrow \mathrm{Sh}_{K_{\mathrm{G}}}(\mathrm{G}).$$

The cycle $z_{K_{\mathrm{H}},K_{\mathrm{G}}} := \delta_{K_{\mathrm{H}},K_{\mathrm{G}},*}[\mathrm{Sh}_{K_{\mathrm{H}}}(\mathrm{H})]$ on $\mathrm{Sh}_{K_{\mathrm{G}}}(\mathrm{G})$ will be called the *special cycle* (for the level $(K_{\mathrm{H}}, K_{\mathrm{G}})$). Very often we choose $K_{\mathrm{H}} = K_{\mathrm{G}} \cap \mathrm{H}(\mathbb{A}_f)$ in which case we simply denote the special cycle by $z_{K_{\mathrm{G}}}$.

Remark 3.3. (i) Note that here our special cycles are different from those appearing in S. S. Kudla, Rapoport, and Yang [2006] and S. Kudla and Rapoport [2014].

(ii) When $\dim_{\mathbb{C}} X_{\mathrm{G}}$ is even, the special cycles are in the middle dimension. When $\dim_{\mathbb{C}} X_{\mathrm{G}}$ is odd, the special cycles are just below the middle dimension, and we will say that they are in the *arithmetic middle dimension* (in the sense that, once extending both Shimura varieties to suitable integral models, we obtain cycles in the middle dimension).

The special cases in the middle dimension are very often related to the study of Tate cycles and automorphic period integrals, e.g., in the pioneering example of Harder, Langlands, and Rapoport [1986], and many of its generalizations.

Below we focus on the case where the special cycles are in the arithmetic middle dimension.

3.1.2 Gross–Zagier pair.

In the case of the Gross–Zagier formula Gross and Zagier [1986], one considers an embedding of an *imaginary* quadratic field F' into $\mathrm{Mat}_{2,\mathbb{Q}}$ (the algebra of 2×2-matrices), and the induced embedding

$$\mathrm{H} = \mathrm{R}_{F'/\mathbb{Q}}\mathbb{G}_m \hookrightarrow \mathrm{G} = \mathrm{GL}_{2,\mathbb{Q}}.$$

Note that $\mathrm{H}_{\mathbb{R}} \simeq \mathbb{C}^\times$ as $\mathbb{R}$-groups (upon a choice of embedding $F' \hookrightarrow \mathbb{C}$). This defines $h_{\mathrm{H}} : \mathbb{S} \to \mathrm{H}_{\mathbb{R}}$, and its composition with the embedding $\mathrm{H}_{\mathbb{R}} \to \mathrm{G}_{\mathbb{R}}$ defines $h_{\mathrm{G}} : \mathbb{S} \to \mathbb{G}_{\mathbb{R}}$. We obtain a special pair $(\mathrm{H}, X_{\mathrm{H}}) \to (\mathrm{G}, X_{\mathrm{G}})$, where

$$\dim X_{\mathrm{G}} = 1, \qquad \dim X_{\mathrm{H}} = 0.$$

In the general case, we replace $F'/\mathbb{Q}$ by a CM extension F'/F of a totally real number field F, and replace $\mathrm{Mat}_{2,\mathbb{Q}}$ by a quaternion algebra B over F that is ramified at all but one archimedean places of F. X. Yuan, S. Zhang, and the author proved Gross–Zagier formula in this generality in Yuan, S.-W. Zhang, and W. Zhang [2013].

3.1.3 Gan–Gross–Prasad pair. Let (H, G) be the Gan–Gross–Prasad pair in Section 2.2.2, but viewed as algebraic groups over $\mathbb{Q}$ (i.e., the pair $(\mathrm{R}_{F/\mathbb{Q}}\mathrm{H}, \mathrm{R}_{F/\mathbb{Q}}\mathrm{G})$). The groups are associated to an embedding $W_{n-1} \subset W_n$ of orthogonal or Hermitian spaces (with respect to F'/F). Now we impose the following conditions.

(i) F is a totally real number field, and in the Hermitian case F'/F is a CM (=totally imaginary quadratic) extension.

(ii) For an archimedean place $\varphi \in \mathrm{Hom}(F, \mathbb{R})$, denote by $\mathrm{sgn}_\varphi(W)$ the signature of $W \otimes_{F,\varphi} \mathbb{R}$ as an orthogonal or Hermitian space over $F' \otimes_{F,\varphi} \mathbb{R}$. Then there exists a distinguished real place $\varphi_0 \in \mathrm{Hom}(F, \mathbb{R})$ such that

$$\mathrm{sgn}_\varphi(W_n) = \begin{cases} (2, n-2), & \varphi = \varphi_0 \\ (0, n), & \varphi \in \mathrm{Hom}(F, \mathbb{R}) \setminus \{\varphi_0\} \end{cases}$$

in the orthogonal case, and

$$\mathrm{sgn}_\varphi(W_n) = \begin{cases} (1, n-1), & \varphi = \varphi_0 \\ (0, n), & \varphi \in \mathrm{Hom}(F, \mathbb{R}) \setminus \{\varphi_0\} \end{cases}$$

in the Hermitian case. In addition, the quotient W_n/W_{n-1} is negative definite at every $\varphi \in \mathrm{Hom}(F, \mathbb{R})$ (so the signature of W_{n-1} is given by similar formulas).

Then Gan, Gross, and Prasad Gan, Gross, and Prasad [2012, §27] prescribe Shimura data that enhance the embedding $\mathrm{H} \hookrightarrow \mathrm{G}$ to a homomorphism of Shimura data $(\mathrm{H}, X_{\mathrm{H}}) \to (\mathrm{G}, X_{\mathrm{G}})$, where the dimensions are

$$\begin{cases} \dim X_{\mathrm{G}} = 2n-5, & \dim X_{\mathrm{H}} = n-3, \quad \text{in the orthogonal case,} \\ \dim X_{\mathrm{G}} = 2n-3, & \dim X_{\mathrm{H}} = n-2, \quad \text{in the Hermitian case.} \end{cases}$$

3.2 The arithmetic Gan–Gross–Prasad conjecture.

3.2.1 Height pairings. Let X be a smooth proper variety over a number field E, and let $\mathrm{Ch}^i(X)$ be the group of codimension-i algebraic cycles on X modulo rational equivalence. We have a cycle class map

$$\mathrm{cl}_i : \mathrm{Ch}^i(X)_{\mathbb{Q}} \longrightarrow H^{2i}(X),$$

where $H^{2i}(X)$ is the Betti cohomology $H^*\big(X(\mathbb{C}), \mathbb{C}\big)$. The kernel is the group of cohomologically trivial cycles, denoted by $\mathrm{Ch}^i(X)_0$.

Conditional on some standard conjectures on algebraic cycles, there is a height pairing defined by Beilinson and Bloch,

$$(3\text{-}1)\qquad (\ ,\ )_{\mathrm{BB}}\colon \mathrm{Ch}^i(X)_{\mathbb{Q},0}\times \mathrm{Ch}^{d+1-i}(X)_{\mathbb{Q},0}\longrightarrow \mathbb{R},\quad d=\dim X.$$

This is unconditionally defined when $i = 1$ (the Néron–Tate height), or when X is an abelian variety Künnemann [2001]. In some situations, cf. Rapoport, Smithling, and W. Zhang [2017b, §6.1], one can define the height pairing unconditionally in terms of the arithmetic intersection theory of Arakelov and Gillet–Soulé Gillet and Soulé [1990, §4.2.10]. This is the case when there exists a smooth proper model $\mathcal{X}$ of X over O_E (this is also true for Deligne–Mumford (DM) stacks X and $\mathcal{X}$).

3.2.2 The arithmetic Gan–Gross–Prasad conjecture. We consider the special cycle in the Gan–Gross–Prasad setting Section 3.1.3, which we also call the *arithmetic diagonal cycle* Rapoport, Smithling, and W. Zhang [2017b]. We will state a version of the arithmetic Gan–Gross–Prasad conjecture assuming some standard conjectures on algebraic cycles (cf. Rapoport, Smithling, and W. Zhang [ibid., §6]), in particular, that we have the height pairing (3-1).

For each $K\subset \mathrm{G}(\mathbb{A}_f)$, one can construct "Hecke–Kunneth" projectors that project the total cohomology of the Shimura variety $\mathrm{Sh}_K(\mathrm{G})$ (or its toroidal compactification) to the odd-degree part (cf. Rapoport, Smithling, and W. Zhang [ibid., §6.2] in the Hermitian case; the same proof works in the orthogonal case). Then we apply this projector to define a cohomologically trivial cycle $z_{K,0}\in \mathrm{Ch}^{n-1}\big(\mathrm{Sh}_K(\mathrm{G})\big)_0$ (with $\mathbb{C}$-coefficient). The classes $\{z_{K,0}\}_{K\subset \mathrm{G}(\mathbb{A}_f)}$ are independent of the choice of our projectors (cf. Rapoport, Smithling, and W. Zhang [ibid., Remark 6.11]), and they form a projective system (with respect to push-forward).

We form the colimit

$$\mathrm{Ch}^{n-1}\big(\mathrm{Sh}(\mathrm{G})\big)_0 := \varinjlim_{K\subset \mathrm{G}(\mathbb{A}_f)} \mathrm{Ch}^{n-1}\big(\mathrm{Sh}_K(\mathrm{G})\big)_0.$$

The height pairing with $\{z_{K,0}\}_{K\subset \mathrm{G}(\mathbb{A}_f)}$ defines a linear functional

$$\mathcal{P}_{\mathrm{Sh}(\mathrm{H})}\colon \mathrm{Ch}^{n-1}\big(\mathrm{Sh}(\mathrm{G})\big)_0 \longrightarrow \mathbb{C}\ .$$

This is the arithmetic version of the automorphic period integral in Section 2.1.2. The group $\mathrm{G}(\mathbb{A}_f)$ acts on the space $\mathrm{Ch}^{n-1}\big(\mathrm{Sh}(\mathrm{G})\big)_0$. For any representation π_f of $\mathrm{G}(\mathbb{A}_f)$, let $\mathrm{Ch}^{n-1}\big(\mathrm{Sh}(\mathrm{G})\big)_0[\pi_f]$ denote the π_f-isotypic component of the Chow group

$$\mathrm{Ch}^{n-1}\big(\mathrm{Sh}(\mathrm{G})\big)_0$$

We are ready to state the arithmetic Gan–Gross–Prasad conjecture Gan, Gross, and Prasad [2012, §27], parallel to Conjecture 2.3.

Conjecture 3.4. *Let π be a tempered cuspidal automorphic representation of* $\mathrm{G}(\mathbb{A})$*, appearing in the cohomology* $H^*(\mathrm{Sh}(\mathrm{G}))$*. The following statements are equivalent.*

(i) *The linear functional $\mathcal{P}_{\mathrm{Sh(H)}}$ does not vanish on the π_f-isotypic component*

$$\mathrm{Ch}^{n-1}\big(\mathrm{Sh}(\mathrm{G})\big)_0[\pi_f].$$

(ii) *The space* $\mathrm{Hom}_{\mathrm{H}(\mathbb{A}_f)}(\pi_f, \mathbb{C}) \neq 0$ *and the first order derivative* $L'(\frac{1}{2}, \pi, R) \neq 0$.

In the orthogonal case with $n \leq 4$, and when the ambient Shimura variety is a curve ($n = 3$), or a product of three curves ($n = 4$), the conjecture is unconditionally formulated. The case $n = 3$ is proved by X. Yuan, S. Zhang, and the author in Yuan, S.-W. Zhang, and W. Zhang [2013]; in fact we proved a refined version. When $n = 4$ and in the triple product case (i.e., the Shimura variety $\mathrm{Sh}_K(\mathrm{G})$ is a product of three curves), X. Yuan, S. Zhang, and the author formulated a refined version of the above conjecture and proved it in some special cases, cf. Yuan, S.-W. Zhang, and W. Zhang [n.d.].

3.2.3 Reformulation in terms of spherical characters. In the Hermitian case, Rapoport, Smithling, and the author in Rapoport, Smithling, and W. Zhang [2017b] stated a version of the arithmetic Gan–Gross–Prasad conjecture that does not depend on standard conjectures on algebraic cycles.

In fact we work with a variant of the Shimura data defined by Gan, Gross, and Prasad Gan, Gross, and Prasad [2012, §27]. We modify the groups $\mathrm{R}_{F/\mathbb{Q}}\mathrm{G}$ and $\mathrm{R}_{F/\mathbb{Q}}\mathrm{H}$ defined previously

$$\begin{aligned}
\mathrm{Z}^{\mathbb{Q}} &:= \mathrm{GU}_1 = \big\{ z \in \mathrm{R}_{F'/\mathbb{Q}}\mathbb{G}_m \bigm| \mathrm{Nm}_{F'/F}(z) \in \mathbb{G}_m \big\}, \\
\widetilde{\mathrm{H}} &:= \mathrm{G}\left(\mathrm{U}_1 \times \mathrm{U}(\mathrm{W_{n-1}})\right) = \big\{ (z,h) \in \mathrm{Z}^{\mathbb{Q}} \times \mathrm{GU}(W_{n-1}) \bigm| \mathrm{Nm}_{F'/F}(z) = c(h) \big\}, \\
\widetilde{\mathrm{G}} &:= \mathrm{G}\left(\mathrm{U}_1 \times \mathrm{U}(\mathrm{W_{n-1}}) \times \mathrm{U}(\mathrm{W_n})\right) \\
&= \big\{ (z,h,g) \in \mathrm{Z}^{\mathbb{Q}} \times \mathrm{GU}(W_{n-1}) \times \mathrm{GU}(W_n) \bigm| \mathrm{Nm}_{F'/F}(z) = c(h) = c(g) \big\},
\end{aligned}$$

where the symbol c denotes the unitary similitude factor. Then we have

$$\widetilde{\mathrm{H}} \xrightarrow{\sim} \mathrm{Z}^{\mathbb{Q}} \times \mathrm{R}_{F/\mathbb{Q}}\mathrm{H}\,, \qquad \widetilde{\mathrm{G}} \xrightarrow{\sim} \mathrm{Z}^{\mathbb{Q}} \times \mathrm{R}_{F/\mathbb{Q}}\mathrm{G}\,. \tag{3-2}$$

We then define natural Shimura data $\big(\widetilde{\mathrm{H}}, \{h_{\widetilde{\mathrm{H}}}\}\big)$ and $\big(\widetilde{\mathrm{G}}, \{h_{\widetilde{\mathrm{G}}}\}\big)$, cf. Rapoport, Smithling, and W. Zhang [2017b, §3]. This variant has the nice feature that the Shimira varieties are of PEL type, i.e., the canonical models are related to moduli problems of abelian varieties with polarizations, endomorphisms, and level structures, cf. Rapoport, Smithling, and W. Zhang [ibid., §4–§5].

For suitable Hermitian spaces and a special level structure $K^{\circ}_{\widetilde{G}} \subset \widetilde{G}(\mathbb{A}_f)$, we can even define *smooth* integral models (over the ring of integers of the reflex field) of the Shimura variety $\mathrm{Sh}_{K^{\circ}_{\widetilde{G}}}(\widetilde{G})$. For a general CM extension F'/F, it is rather involved to state this level structure Rapoport, Smithling, and W. Zhang [ibid., Remark 6.19] and define the integral models Rapoport, Smithling, and W. Zhang [ibid., §5]. For simplicity, from now on we consider a special case, when $F = \mathbb{Q}$ and $F' = F[\varpi]$ is an imaginary quadratic field. We further assume that the prime 2 is split in F'. We choose $\varpi \in F'$ such that $(\varpi) \subset O_{F'}$ is the product of all ramified prime ideals in $O_{F'}$.

We first define an auxiliary moduli functor $\mathcal{M}_{(r,s)}$ over $\operatorname{Spec} O_{F'}$ for $r + s = n$ (similar to S. Kudla and Rapoport [2014, §13.1]). For a locally noetherian scheme S over $\operatorname{Spec} O_{F'}$, $\mathcal{M}_{(r,s)}(S)$ is the groupoid of triples (A, ι, λ) where

- (A, ι) is an abelian scheme over S, with $O_{F'}$-action $\iota : O_{F'} \to \mathrm{End}(A)$ satisfying the Kottwitz condition of signature (r, s), and

- $\lambda : A \to A^{\vee}$ is a polarization whose Rosati involution induces on $O_{F'}$ the non-trivial Galois automorphism of F'/F, and such that $\ker(\lambda)$ is contained in $A[\iota(\varpi)]$ of rank $\#(O_{F'}/(\varpi))^n$ (resp. $\#(O_{F'}/(\varpi))^{n-1}$) when $n = r + s$ is even (resp. odd). In particular, we have $\ker(\lambda) = A[\iota(\varpi)]$ if $n = r + s$ is even.

Now we assume that $(r, s) = (1, n-1)$ or $(n-1, 1)$. We further impose the *wedge condition* and the *(refined) spin condition*, cf. Rapoport, Smithling, and W. Zhang [2017b, §4.4]. The functor is represented by a Deligne–Mumford stack again denoted by $\mathcal{M}_{(r,s)}$. It is *smooth* over $\operatorname{Spec} O_{F'}$, despite the ramification of the field extension $F'/\mathbb{Q}$, cf. Rapoport, Smithling, and W. Zhang [ibid., §4.4]. Then we have an integral model of copies of the Shimura variety $\mathrm{Sh}_{K^{\circ}_{\widetilde{G}}}(\widetilde{G})$ defined by

$$\mathcal{M}_{K^{\circ}_{\widetilde{G}}}(\widetilde{G}) = \mathcal{M}_{(0,1)} \times_{\operatorname{Spec} O_{F'}} \mathcal{M}_{(1,n-2)} \times_{\operatorname{Spec} O_{F'}} \mathcal{M}_{(1,n-1)}.$$

(In Rapoport, Smithling, and W. Zhang [ibid., §5.1] we do cut out the desired Shimura variety with the help of a sign invariant. Here, implicitly we need to replace this space by its toroidal compactification.)

We now describe the arithmetic diagonal cycle (or rather, its integral model) for the level $K^{\circ}_{\widetilde{H}} = K^{\circ}_{\widetilde{G}} \cap \widetilde{H}(\mathbb{A}_f)$. When n is odd (so $n-1$ is even), we define

$$\mathcal{M}_{K^{\circ}_{\widetilde{H}}}(\widetilde{H}) = \mathcal{M}_{(0,1)} \times_{\operatorname{Spec} O_{F'}} \mathcal{M}_{(1,n-2)},$$

and we can define an embedding explicitly by "taking products" (one sees easily that the conditions on the kernels of polarizations are satisfied):

(3-3)
$$\begin{array}{ccc} \mathfrak{M}_{K^{\circ}_{\widetilde{\mathrm{H}}}}(\widetilde{\mathrm{H}}) & \longrightarrow & \mathfrak{M}_{K^{\circ}_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}}) \\ (A_0, \iota_0, \lambda_0, A^\flat, \iota^\flat, \lambda^\flat) & \longmapsto & (A_0, \iota_0, \lambda_0, A^\flat, \iota^\flat, \lambda^\flat, A^\flat \times A_0, \iota^\flat \times \iota_0, \lambda^\flat \times \lambda_0). \end{array}$$

When n is even, the situation is more subtle; see Rapoport, Smithling, and W. Zhang [2017b, §4.4].

With the smooth integral model, we have an unconditionally defined height pairing (3-1) on $X = \mathrm{Sh}_{K^{\circ}_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$. Now we again apply a suitable Hecke–Kunneth projector to the cycle z_K for $K = K^{\circ}_{\widetilde{\mathrm{G}}}$, and we obtain a cohomologically trivial cycle $z_{K,0} \in \mathrm{Ch}(\mathrm{Sh}_{K^{\circ}_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}}))_0$. We define

(3-4)
$$\mathrm{Int}(f) = \Big(R(f) * z_{K,0},\ z_{K,0}\Big)_{\mathrm{BB}}, \quad f \in \mathcal{H}\left(\widetilde{\mathrm{G}}, K^{\circ}_{\widetilde{\mathrm{G}}}\right),$$

where $R(f)$ is the associated Hecke correspondence. Let $\mathcal{H}^{\mathrm{ram}_{F'}}(\widetilde{\mathrm{G}})$ be the spherical Hecke algebra away from the set $\mathrm{ram}_{F'}$ of primes ramified in F'/F.

Parallel to Conjecture 2.11, we can state an alternative version of the arithmetic Gan–Gross–Prasad conjecture in terms of spherical characters for the special level $K^{\circ}_{\widetilde{\mathrm{G}}}$.

Conjecture 3.5. *There is a decomposition*

$$\mathrm{Int}(f) = \sum_{\pi} \mathrm{Int}_\pi(f), \quad \textit{for all } f \in \mathcal{H}\left(\widetilde{\mathrm{G}}, K^{\circ}_{\widetilde{\mathrm{G}}}\right),$$

where the sum runs over all automorphic representations of $\widetilde{\mathrm{G}}(\mathbb{A})$ *that appear in the cohomology* $H^*(\mathrm{Sh}_{\widetilde{G},K_{\mathrm{G}}})$ *and are trivial on* $Z^{\mathbb{Q}}(\mathbb{A})$, *and* Int_π *is an eigen-distribution for the spherical Hecke algebra* $\mathcal{H}^{\mathrm{ram}_{F'}}(\widetilde{\mathrm{G}})$ *with eigen-character* $\lambda_{\pi^{\mathrm{ram}_{F'}}}$ *in the sense of* (2-9).

If such a representation π *is tempered, then*

$$\mathrm{Int}_\pi(f) = 2^{-\beta_\pi} \mathcal{L}'(1/2, \pi) \prod_{v<\infty} \mathbb{I}_{\pi_v}(f_v).$$

Here the constant β_π is the same as in Section 2.2.4, and there is a natural extension of the local spherical characters $\mathbb{I}_{\pi_v}$ to the triple $(\widetilde{\mathrm{G}}, \widetilde{\mathrm{H}}, \widetilde{\mathrm{H}})$.

Remark 3.6. Conjecture 3.5 can be viewed as a refined version of the arithmetic Gan–Gross–Prasad Conjecture 3.4. Other refinements were also given in by the author W. Zhang [2009] and independently by S. S. Zhang [2010]. Both of them rely on standard conjectures on height pairings, and hence are conditional.

4 Shtukas and higher Gross–Zagier formula

Now let F be the function field of a geometrically connected smooth proper curve X over a finite field $k = \mathbb{F}_q$. We may consider the analog of the special pair of Shimura data (cf. 3.1.1) in the context of Shtukas. Now there is much more freedom since we do not have the restriction from the archimedean place. One may choose an r-tuple of coweights of G to define G–Shtukas (with r-modifications), and the resulting moduli space lives over the r-fold power

$$X^r = \underbrace{X \times_{\operatorname{Spec} k} \dots \times_{\operatorname{Spec} k} X}_{r \text{ times}}.$$

This feature is completely missing in the number field case, where we only have two available options:

(i) when $r = 0$, the automorphic quotient $[\mathrm{G}]_K$ plays an analogous role, cf. 2.1.1.

(ii) when $r = 1$, we have Shimura varieties $\mathrm{Sh}_{K_\mathrm{G}}(\mathrm{G})$ associated to a Shimura datum $(\mathrm{G}, \{h_\mathrm{G}\})$. These varieties live over Spec E for a number field E.

In Yun and W. Zhang [2017] Yun and the author studied a simplest case, i.e., the special cycle on the moduli stack of rank two Shtukas with arbitrary number r of modifications. We connect their intersection numbers to the r-th order derivative of certain L-functions. We may view the result as an analog of Waldspurger's formula (for $r = 0$) and the Gross–Zagier formula (for $r = 1$).

4.1 The Heegner–Drinfeld cycle. Let $\mathrm{G} = \mathrm{PGL}_2$ and let Bun_2 be the stack of rank two vector bundles on X. The Picard stack Pic_X acts on Bun_2 by tensoring the line bundle. Then $\mathrm{Bun}_\mathrm{G} = \mathrm{Bun}_2 / \mathrm{Pic}_X$ is the moduli stack of G-torsors over X.

Let r be an *even* integer. Let $\mu \in \{\pm\}^r$ be an r-tuple of signs such that exactly half of them are equal to $+$. Let Hk_2^μ be the Hecke stack, i.e., $\mathrm{Hk}_2^\mu(S)$ is the groupoid of data

$$(\mathcal{E}_0, \cdots, \mathcal{E}_r, x_1, \cdots, x_r, f_1, \cdots, f_r),$$

where the $\mathcal{E}_i$'s are vector bundles of rank two over $X \times S$, the x_i's are S-points of X, each f_i is a minimal upper (i.e., increasing) modification if $\mu_i = +$, and minimal lower (i.e., decreasing) modification if $\mu_i = -$, and the i-th modification takes place along the graph of $x_i : S \to X$,

$$\mathcal{E}_0 \overset{f_1}{\dashrightarrow} \mathcal{E}_1 \overset{f_2}{\dashrightarrow} \cdots \overset{f_r}{\dashrightarrow} \mathcal{E}_r .$$

The Picard stack Pic_X acts on Hk_2^μ by simultaneously tensoring the line bundle. Define $\mathrm{Hk}_\mathrm{G}^\mu = \mathrm{Hk}_2^\mu / \mathrm{Pic}_X$. Recording $\mathcal{E}_i$ defines a projection $p_i : \mathrm{Hk}_\mathrm{G}^\mu \to \mathrm{Bun}_\mathrm{G}$.

The moduli stack $\mathrm{Sht}_{\mathrm{G}}^{\mu}$ of Drinfeld G-Shtukas of type μ for the group G is defined by the cartesian diagram

$$\begin{array}{ccc} \mathrm{Sht}_{\mathrm{G}}^{\mu} & \longrightarrow & \mathrm{Hk}_{\mathrm{G}}^{\mu} \\ \downarrow & & \downarrow {\scriptstyle (p_0, p_r)} \\ \mathrm{Bun}_{\mathrm{G}} & \xrightarrow{(\mathrm{id},\mathrm{Fr})} & \mathrm{Bun}_{\mathrm{G}} \times \mathrm{Bun}_{\mathrm{G}} \end{array} . \tag{4-1}$$

The stack $\mathrm{Sht}_{\mathrm{G}}^{\mu}$ is a Deligne–Mumford stack over X^r, and the natural morphism

$$\pi_{\mathrm{G}}^{\mu} : \mathrm{Sht}_{\mathrm{G}}^{\mu} \longrightarrow X^r$$

is smooth of relative dimension r.

Let $\nu : X' \to X$ be a finite étale cover of degree 2 such that X' is also geometrically connected. Denote by $F' = k(X')$ the function field. Let $\mathrm{T} = (\mathrm{R}_{F'/F}\mathbb{G}_m)/\mathbb{G}_m$ be the non-split torus associated to the double cover X' of X. The stack $\mathrm{Sht}_{\mathrm{T}}^{\mu}$ of T-Shtukas is defined analogously, with the rank two bundles $\mathcal{E}_i$ replaced by line bundles $\mathcal{L}_i$ on X', and the points x_i on X'. Then we have a map

$$\pi_{\mathrm{T}}^{\mu} : \mathrm{Sht}_{\mathrm{T}}^{\mu} \longrightarrow X'^r$$

which is a torsor under the finite Picard stack $\mathrm{Pic}_{X'}(k)/\mathrm{Pic}_X(k)$. In particular, $\mathrm{Sht}_{\mathrm{T}}^{\mu}$ is a proper smooth Deligne–Mumford stack over $\mathrm{Spec}\, k$.

There is a natural finite morphism of stacks over X^r, induced by the natural map $\nu_* : \mathrm{Pic}_{X'} \to \mathrm{Bun}_2$,

$$\mathrm{Sht}_{\mathrm{T}}^{\mu} \longrightarrow \mathrm{Sht}_{\mathrm{G}}^{\mu} .$$

It induces a finite morphism

$$\theta^{\mu} : \mathrm{Sht}_{\mathrm{T}}^{\mu} \longrightarrow \mathrm{Sht}_{\mathrm{G}}'^{\mu} := \mathrm{Sht}_{\mathrm{G}}^{\mu} \times_{X^r} X'^r .$$

This defines a class in the Chow group of proper cycles of dimension r with $\mathbb{Q}$-coefficients,

$$Z_{\mathrm{T}}^{\mu} := \theta_*^{\mu}[\mathrm{Sht}_{\mathrm{T}}^{\mu}] \in \mathrm{Ch}_{c,r}(\mathrm{Sht}_{\mathrm{G}}'^{\mu})_{\mathbb{Q}}. \tag{4-2}$$

In analogy to the Heegner cycle in the Gross–Zagier formula Gross and Zagier [1986] and Yuan, S.-W. Zhang, and W. Zhang [2013] in the number field case, we call Z_{T}^{μ} the *Heegner–Drinfeld cycle* in our setting.

Remark 4.1. The construction of the Heegner–Drinfeld cycle extends naturally to higher rank Shtukas (of rank n over X', respectively rank $2n$ over X) of type $\mu = (\mu_1, \cdots, \mu_r)$. Here μ_i are coweights of GL_n (or GL_{2n}) given by $(\pm 1, 0, \cdots, 0) \in \mathbb{Z}^n$ (or $\mathbb{Z}^{2n}$).

4.2 Taylor expansion of L-functions. Consider the middle degree cohomology with compact support

$$V = \mathrm{H}_c^{2r}((\mathrm{Sht}'^{\mu}_G) \otimes_k \overline{k}, \overline{\mathbb{Q}}_\ell)(r).$$

This vector space is endowed with the cup product

$$(\cdot,\cdot) : V \times V \longrightarrow \overline{\mathbb{Q}}_\ell \ .$$

Let ℓ be a prime number different from p. Let $K = \prod_v \mathrm{G}(O_{F_v})$, and let $\mathscr{H}_{\overline{\mathbb{Q}}_\ell} = \mathscr{H}(\mathrm{G}(\mathbb{A}), K)$ be the spherical Hecke algebra with $\overline{\mathbb{Q}}_\ell$-coefficients. For any maximal ideal $\mathfrak{m} \subset \mathscr{H}_{\overline{\mathbb{Q}}_\ell}$, we define the generalized eigenspace of V with respect to $\mathfrak{m}$ by

$$V_{\mathfrak{m}} = \cup_{i>0} V[\mathfrak{m}^i].$$

We also define a subspace V_{Eis} with the help of an *Eisenstein ideal*, cf Yun and W. Zhang [2017, §4.1.2]. Then we prove that there is a spectral decomposition, i.e., an orthogonal decomposition of $\mathscr{H}_{\overline{\mathbb{Q}}_\ell}$-modules,

$$V = V_{\mathrm{Eis}} \oplus \left(\bigoplus_{\mathfrak{m}} V_{\mathfrak{m}} \right), \tag{4-3}$$

where $\mathfrak{m}$ runs over a finite set of maximal ideals of $\mathscr{H}_{\overline{\mathbb{Q}}_\ell}$, and each $V_{\mathfrak{m}}$ is an $\mathscr{H}_{\overline{\mathbb{Q}}_\ell}$-module of finite dimension over $\overline{\mathbb{Q}}_\ell$ supported at the maximal ideal $\mathfrak{m}$; see Yun and W. Zhang [ibid., Thm. 7.16] for a more precise statement.

Let π be an everywhere unramified cuspidal automorphic representation π of $\mathrm{G}(\mathbb{A})$. The standard L-function $L(\pi, s)$ is a polynomial of degree $4(g-1)$ in $q^{-s-1/2}$, where g is the genus of X. Let $\pi_{F'}$ be the base change to F', and let $L(\pi_{F'}, s)$ be its standard L-function. Let $L(\pi, \mathrm{Ad}, s)$ be the adjoint L-function of π and define

$$\mathscr{L}(\pi_{F'}, s) = \epsilon(\pi_{F'}, s)^{-1/2} \frac{L(\pi_{F'}, s)}{L(\pi, \mathrm{Ad}, 1)}, \tag{4-4}$$

where the the square root is understood as $\epsilon(\pi_{F'}, s)^{-1/2} = q^{4(g-1)(s-1/2)}$. In particular, we have a functional equation:

$$\mathscr{L}(\pi_{F'}, s) = \mathscr{L}(\pi_{F'}, 1-s).$$

We consider the Taylor expansion at the central point $s = 1/2$:

$$\mathscr{L}(\pi_{F'}, s) = \sum_{r \geq 0} \mathscr{L}^{(r)}(\pi_{F'}, 1/2) \frac{(s-1/2)^r}{r!},$$

i.e.,

$$\mathcal{L}^{(r)}(\pi_{F'}, 1/2) = \frac{d^r}{ds^r}\Big|_{s=0} \left(\epsilon(\pi_{F'}, s)^{-1/2} \frac{L(\pi_{F'}, s)}{L(\pi, \mathrm{Ad}, 1)} \right).$$

If r is odd, by the functional equation we have

$$\mathcal{L}^{(r)}(\pi_{F'}, 1/2) = 0.$$

Now we fix an isomorphism $\mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$. Let $\mathfrak{m} = \mathfrak{m}_\pi$ be the kernel of the associated character $\lambda_\pi : \mathcal{H}_{\overline{\mathbb{Q}}_\ell} \to \overline{\mathbb{Q}}_\ell$, and rename $V_\mathfrak{m}$ in (4-3) as V_π. Then our main result in Yun and W. Zhang [2017] relates the r-th Taylor coefficient to the self-intersection number of the π-component of the Heegner–Drinfeld cycle $\theta^\mu_*[\mathrm{Sht}^\mu_\mathrm{T}]$.

Theorem 4.2. *Let π be an everywhere unramified cuspidal automorphic representation of $\mathrm{G}(\mathbb{A})$. Let $[\mathrm{Sht}^\mu_\mathrm{T}]_\pi \in V_\pi$ be the projection of the cycle class of $\mathrm{cl}\left(\theta^\mu_*[\mathrm{Sht}^\mu_\mathrm{T}]\right) \in V$ to the direct summand V_π under the decomposition (4-3). Then*

$$\frac{1}{2(\log q)^r} |\omega_X| \, \mathcal{L}^{(r)}(\pi_{F'}, 1/2) = \left([\mathrm{Sht}^\mu_\mathrm{T}]_\pi, \quad [\mathrm{Sht}^\mu_T]_\pi \right),$$

where ω_X is the canonical divisor, and $|\omega_X| = q^{-2g+2}$.

Remark 4.3. Here we only consider étale double covers X'/X, and everywhere unramified π (whence the L-function has nonzero Taylor coefficients in even degrees only). In Yun and W. Zhang [n.d.], Yun and the author are extending the theorem above to the case when X'/X is ramified at a finite set R and π has Iwahori levels at Σ such that $R \cap \Sigma = \varnothing$.

4.3 Comparison with the conjecture of Birch and Swinnerton-Dyer. Let π be as in Theorem 4.2, and ρ_π the associated local system of rank two over the curve X by the global Langlands correspondence. Let $W'_\pi = H^1\left(X' \times \bar{k},\ \rho_\pi\right)$, a $\overline{\mathbb{Q}}_\ell$-vector space with the Frobenius endomorphism Fr. The L-function $L(\pi_{F'}, s)$ is then given by

$$L(\pi_{F'}, s - 1/2) = \det\left(1 - q^{-s}\,\mathrm{Fr} \,\middle|\, W'_\pi\right).$$

In particular, the dimension of the eigenspace $W'^{\,\mathrm{Fr}=q}_\pi$ is at most $\mathrm{ord}_{s=1/2} L(\pi_{F'}, s)$ (the conjectural semi-simplicity of Fr implies an equality). It is *expected* that, the complex $\mathbf{R}\pi^\mu_{\mathrm{G},!}\overline{\mathbb{Q}}_\ell$ on X^r decomposes as a direct sum of $\mathcal{H}_{\overline{\mathbb{Q}}_\ell}$-modules

$$\mathbf{R}\pi^\mu_{\mathrm{G},!}\overline{\mathbb{Q}}_\ell = \Big(\bigoplus_{\pi \text{ cuspidal}} \pi^K \otimes (\underbrace{\rho_\pi \boxtimes \cdots \boxtimes \rho_\pi}_{r \quad \text{times}}) \Big) \bigoplus \text{"a direct summand"},$$

such that $V_\pi = V_{\mathfrak{m}_\pi}$ in (4-3) (for $\mathfrak{m}_\pi = \ker(\lambda_\pi)$) corresponds to $\pi^K \otimes W'^{\otimes r}_\pi$. From now on we assume this decomposition. Then the cohomology class of the Heegner–Drinfeld cycle defines an element $Z_\pi^\mu \in \pi^K \otimes W_\pi'^{\otimes r}$. One can show that Z_π^μ is an eigen-vector for the operator $\mathrm{id} \otimes \mathrm{Fr}^{\otimes r}$ with eigenvalue q^r. Then Theorem 4.2 shows that this class does not vanish when $r \geq \mathrm{ord}_{s=1/2} L(\pi_{F'}, s)$, provided that $L(\pi_{F'}, s)$ is not a constant.

Conjecture 4.4. *Let* $r = \mathrm{ord}_{s=1/2} L(\pi_{F'}, s)$. *The class* Z_π^μ *belongs to*

$$\pi^K \otimes \wedge^r \left(W_\pi'^{\mathrm{Fr}=q}\right).$$

Note that the generalization of the conjecture of Birch and Swinnerton-Dyer to function fields by Artin and Tate predicts that $\dim W_\pi'^{\mathrm{Fr}=q} = \mathrm{ord}_{s=1/2} L(\pi_{F'}, s)$.

We have a similar conjecture when X'/X is ramified at a finite set R and π has Iwahori levels at Σ such that $R \cap \Sigma = \varnothing$, cf. Yun and W. Zhang [ibid.]. In a forthcoming work, Yun and the author plan to prove that

(i) Let $r_0 \geq 0$ be the smallest integer r such that $Z_\pi^\mu \neq 0$ for some $\mu \in \{\pm\}^r$. Then $\dim W_\pi'^{\mathrm{Fr}=q} = r_0$, and the class Z_π^μ gives a basis to the line $\pi^K \otimes \wedge^r \left(W_\pi'^{\mathrm{Fr}=q}\right)$.

(ii) $\mathrm{ord}_{s=1/2} L(\pi_{F'}, s) = 1$ *if and only if* $\dim W_\pi'^{\mathrm{Fr}=q} = 1$. In particular, if $\mathrm{ord}_{s=1/2} L(\pi_{F'}, s) = 3$, then $\dim W_\pi'^{\mathrm{Fr}=q} = 3$.

5 Relative trace formula

5.1 An overview of RTF. A natural tool to study automorphic period integrals is the *relative trace formula* (RTF) introduced by Jacquet. For the reader's convenience, we give a very brief overview of the relative trace formula (cf. the survey articles Jacquet [2005], Lapid [2006, 2010], and Offen [2009]).

We start with a triple $(\mathrm{G}, \mathrm{H}_1, \mathrm{H}_2)$ consisting of a reductive group G and two subgroups $\mathrm{H}_1, \mathrm{H}_2$ defined over F. Known examples suggest that we may further assume that the pairs $(\mathrm{H}_i, \mathrm{G})$ are spherical (cf. Section 2.1.3), although this is not essential to our informal discussion here.

To a test function $f \in \mathcal{C}_c^\infty(\mathrm{G}(\mathbb{A}))$ we associate an automorphic kernel function,

$$K_f(x, y) := \sum_{\gamma \in \mathrm{G}(F)} f(x^{-1}\gamma y), \quad x, y \in \mathrm{G}(\mathbb{A}),$$

which is invariant under $\mathrm{G}(F)$ for both variables x and y. This defines an integral operator representing $R(f)$ for the action R of $\mathrm{G}(\mathbb{A})$ on the Hilbert space $L^2([\mathrm{G}])$. Therefore the kernel function has a spectral decomposition, and the contribution of a cuspidal automorphic representation π to the kernel function is given by

$$K_{\pi,f}(x, y) = \sum_{\varphi \in \mathrm{OB}(\pi)} \big(\pi(f)\varphi\big)(x)\, \overline{\varphi(y)}, \tag{5-1}$$

where the sum runs over an orthonormal basis $\mathrm{OB}(\pi)$ of π (with respect to the Petersson inner product).

Then we consider a linear functional on $\mathcal{C}_c^\infty(\mathrm{G}(\mathbb{A}))$,

$$\mathbb{I}(f) = \int_{[\mathrm{H}_1]} \int_{[\mathrm{H}_2]} K_f(h_1, h_2)\, dh_1\, dh_2. \tag{5-2}$$

The spectral contribution (5-1) from an automorphic representation π is the (global) *spherical character* (relative to $(\mathrm{H}_1, \mathrm{H}_2)$), denoted by $\mathbb{I}_\pi(f)$. Similar to (2-8), this is equal to

$$\mathbb{I}_\pi(f) = \sum_{\phi \in \mathrm{OB}(\pi)} \mathcal{P}_{\mathrm{H}_1}(\pi(f)\phi)\overline{\mathcal{P}_{\mathrm{H}_2}(\phi)}.$$

Let $\mathrm{H}_{1,2} := \mathrm{H}_1 \times \mathrm{H}_2$. Then $\mathrm{H}_{1,2}$ acts on G by $(h_1, h_2)\colon \gamma \mapsto h_1^{-1}\gamma h_2$. For certain nice orbits $\gamma \in \mathrm{G}(F)/\mathrm{H}_{1,2}(F)$, we can define orbital integrals (relative to $\mathrm{H}_{1,2}$):

$$\mathrm{Orb}(\gamma, f) := \mathrm{vol}([\mathrm{H}_{1,2,\gamma}]) \int_{\mathrm{H}_{1,2}(\mathbb{A})/\mathrm{H}_{1,2,\gamma}(\mathbb{A})} f(h_1^{-1}\gamma h_2)\, dh_1\, dh_2, \tag{5-3}$$

where $\mathrm{H}_{1,2,\gamma}$ denotes the stabilizer of γ, and vol stands for "volume".

The relative trace formula attached to the triple $(\mathrm{G}, \mathrm{H}_1, \mathrm{H}_2)$ is then the identity between the spectral expansion and the geometric expansion of $\mathbb{I}(f)$:

$$\sum_\pi \mathbb{I}_\pi(f) + \cdots = \sum_\gamma \mathrm{Orb}(\gamma, f) + \cdots,$$

where the $\cdots$ parts need more care (in fact saying so is oversimplifying). We will use $\mathrm{RTF}_{(\mathrm{G},\mathrm{H}_1,\mathrm{H}_2)}$ to stand for the above relative trace formula identity. This is only a very coarse form, and depending on the triple $(\mathrm{G}, \mathrm{H}_1, \mathrm{H}_2)$ the identity may need further refinements such as stabilization, as experience with the Arthur–Selberg trace formula suggests.

Remark 5.1. When we take the triple $(\mathrm{H} \times \mathrm{H}, \Delta_\mathrm{H}, \Delta_\mathrm{H})$, where $\Delta_\mathrm{H} \subset \mathrm{H} \times \mathrm{H}$ is the diagonal embedding of H, the associated relative trace formula is essentially equivalent to the Arthur–Selberg trace formula associated to H. Therefore the RTF can be viewed as a generalization of the Arthur–Selberg trace formula.

In application to questions such as the Gan–Gross–Prasad conjecture, we need to compare two RTFs that are close to each other,

$$\mathrm{RTF}_{(\mathrm{G},\mathrm{H}_1,\mathrm{H}_2)} \longleftrightarrow \mathrm{RTF}_{(\mathrm{G}',\mathrm{H}_1',\mathrm{H}_2')}.$$

The comparison allows us to connect the automorphic periods on G to those on G′. There are many successful examples, although it is a subtle question how to seek comparable RTFs in general.

5.2 Jacquet–Rallis RTFs.

5.2.1 . We recall the two RTFs constructed by Jacquet and Rallis [2011] to attack the Gan–Gross–Prasad conjecture in the Hermitian case (cf. Section 2.2.2).

The first RTF deals with the automorphic H-period integral on G and is associated to the triple $(\mathrm{G}, \mathrm{H}, \mathrm{H})$. The second one is associated to the triple $(\mathrm{G}', \mathrm{H}_1', \mathrm{H}_2')$ where

$$\mathrm{G}' = \mathrm{R}_{F'/F}(\mathrm{GL}_{n-1} \times \mathrm{GL}_n),$$

and

$$\mathrm{H}_1' = \mathrm{R}_{F'/F}\mathrm{GL}_{n-1}, \quad \mathrm{H}_2' = \mathrm{GL}_{n-1} \times \mathrm{GL}_n,$$

where $(\mathrm{H}_1', \mathrm{G}')$ is the Rankin–Selberg pair, and $(\mathrm{H}_2', \mathrm{G}')$ the Flicker–Rallis pair, cf. Section 2.1.3. Moreover it is necessary to insert a quadratic character of $\mathrm{H}_2'(\mathbb{A})$:

$$\eta = \eta_{n-1,n} : (h_{n-1}, h_n) \in \mathrm{H}_2'(\mathbb{A}) \longmapsto \eta_{F'/F}^{n-2}(\det(h_{n-1}))\eta_{F'/F}^{n-1}(\det(h_n)),$$

where $\eta_{F'/F} : F^\times\backslash\mathbb{A}^\times \to \{\pm 1\}$ is the quadratic character associated to F'/F by class field theory.

For the later application to the arithmetic Gan–Gross–Prasad conjecture, we introduce (cf. W. Zhang [2012b, §3.1]) the global distribution on $\mathrm{G}'(\mathbb{A})$ parameterized by a complex variable $s \in \mathbb{C}$,

(5-4)
$$\mathbb{J}(f', s) = \int_{[\mathrm{H}_1']}\int_{[\mathrm{H}_2']} K_{f'}(h_1, h_2)\,\left|\det(h_1)\right|^s \eta(h_2)\, dh_1\, dh_2, \quad f' \in \mathcal{C}_c^\infty(\mathrm{G}'(\mathbb{A})).$$

We set

$$\mathbb{J}(f') = \mathbb{J}(f', 0).$$

Remark 5.2. Due to the presence of the Flicker–Rallis pair $(\mathrm{H}_2', \mathrm{G}')$, the cuspidal part of the spectral side in $\mathrm{RTF}_{(\mathrm{G}',\mathrm{H}_1',\mathrm{H}_2')}$ only contains those automorphic representations that are in the image of the quadratic base change from unitary groups. This gives the hope that the spectral sides of the two RTFs should match.

Remark 5.3. In Gan, Gross, and Prasad [2012], Gan, Gross, and Prasad also made global conjectures for $\mathrm{SO}_n \times \mathrm{SO}_m$ and $\mathrm{U}_n \times \mathrm{U}_m$ when $|n-m| > 1$. Towards them in the Hermitian cases, Y. Liu in Liu [2014, 2016] has generalized the construction of Jacquet and Rallis.

5.2.2 Geometric terms: orbital integrals. In the comparison of geometric sides of two RTFs, we need to match orbits and orbital integrals. We review the comparison in the Jacquet–Rallis case.

We call an element $\gamma \in \mathrm{G}(F)$ *regular semi-simple* (relative to the action of $\mathrm{H}_{1,2} = \mathrm{H}_1 \times \mathrm{H}_2$) if its orbit under $\mathrm{H}_{1,2}$ is Zariski closed and its stabilizer is of minimal dimension. The regular semi-simple orbits will be the nice ones for the study of orbital integrals. For the triples $(\mathrm{G}, \mathrm{H}, \mathrm{H})$ and $(\mathrm{G}', \mathrm{H}'_1, \mathrm{H}'_2)$ in the Jacquet–Rallis RTFs, the condition is equivalent to γ having Zariski closed orbit and trivial stabilizer. In particular, for such γ the orbital integral (5-3) simplifies. We denote by $\mathrm{G}(F)_{\mathrm{rs}}$ (resp. $\mathrm{G}'(F)_{\mathrm{rs}}$) the set of regular semi-simple elements in $\mathrm{G}(F)$ (resp. $\mathrm{G}'(F)$). We denote by $[\mathrm{G}(F)_{\mathrm{rs}}]$ and $[\mathrm{G}'(F)_{\mathrm{rs}}]$ the respective sets of orbits.

Depending on the pair of Hermitian spaces $W := (W_{n-1}, W_n)$, we denote the triple $(\mathrm{G}, \mathrm{H}, \mathrm{H})$ by $(\mathrm{G}_W, \mathrm{H}_W, \mathrm{H}_W)$. We consider the equivalence relation $(W_{n-1}, W_n) \sim (W'_{n-1}, W'_n)$ if there is a scalar $\lambda \in F^\times$ such that we have isometries $W'_{n-1} \simeq {}^\lambda W_{n-1}$ and $W'_n \simeq {}^\lambda W_n$, where the left superscript changes the Hermitian form by a multiple λ. There is a natural bijection (cf. W. Zhang [2012b, §2], Rapoport, Smithling, and W. Zhang [2017a, §2])

$$\coprod_W [\mathrm{G}_W(F)_{\mathrm{rs}}] \xrightarrow{\sim} [\mathrm{G}'(F)_{\mathrm{rs}}], \tag{5-5}$$

where the left hand side runs over all pairs $W = (W_{n-1}, W_n)$ up to equivalence. This bijection holds for any quadratic extension of fields F'/F of characteristic not equal to 2.

Now we let F'/F be a quadratic extension of *local fields*. For $g \in \mathrm{G}_W(F)_{\mathrm{rs}}$ and $f \in \mathcal{C}_c^\infty(\mathrm{G}_W(F))$ we introduce the orbital integral

$$\mathrm{Orb}(g, f) = \int_{\mathrm{H}_{1,2}(F)} f(h_1^{-1} g h_2)\, dh_1\, dh_2. \tag{5-6}$$

For $\gamma \in \mathrm{G}'(F)_{\mathrm{rs}}$, $f' \in \mathcal{C}_c^\infty(\mathrm{G}'(F))$, and $s \in \mathbb{C}$, we introduce the (weighted) orbital integral

$$\mathrm{Orb}(\gamma, f', s) = \int_{\mathrm{H}'_{1,2}(F)} f(h_1^{-1} \gamma h_2) \lvert \det(h_1) \rvert^s \eta(h_2)\, dh_1\, dh_2. \tag{5-7}$$

We set

$$\mathrm{Orb}(\gamma, f') := \mathrm{Orb}(\gamma, f', 0) \quad \text{and} \quad \partial\mathrm{Orb}(\gamma, f') := \frac{d}{ds}\Big|_{s=0} \mathrm{Orb}(\gamma, f', s). \tag{5-8}$$

Definition 5.4. (i) We say that $g \in \mathrm{G}_W(F)_{\mathrm{rs}}$ and $\gamma \in \mathrm{G}'(F)_{\mathrm{rs}}$ *match* if their orbits correspond to each other under (5-5).

(ii) Dually, we say that a function $f' \in \mathcal{C}_c^\infty(\mathrm{G}'(F))$ and a tuple $\{f_W \in \mathcal{C}_c^\infty(\mathrm{G}_W(F))\}$ indexed by W (up to equivalence) are *transfers* of each other if for each W and each $g \in \mathrm{G}_W(F)_{\mathrm{rs}}$,

$$\mathrm{Orb}(g, f_W) = \omega(\gamma)\,\mathrm{Orb}(\gamma, f')$$

whenever $\gamma \in \mathrm{G}'(F)_{\mathrm{rs}}$ matches g. Here $\omega(\gamma)$ is a certain explicit transfer factor W. Zhang [2014b] and Rapoport, Smithling, and W. Zhang [2018].

(iii) We say that a component f_W in the tuple is a *transfer* of f' if the remaining components of the tuple are all zero.

In W. Zhang [2014b] we prove the following.

Theorem 5.5. *Let F'/F be a quadratic extension of p-adic local fields (then there are two equivalence classes of pairs of Hermitian spaces denoted by $W, W^\flat$ respectively). Then for any $f' \in \mathcal{C}_c^\infty(\mathrm{G}'(F))$ there exists a transfer $(f_0, f_1) \in \mathcal{C}_c^\infty(\mathrm{G}_W(F)) \times \mathcal{C}_c^\infty(\mathrm{G}_{W^\flat}(F))$, and for any pair $(f_0, f_1) \in \mathcal{C}_c^\infty(\mathrm{G}_W(F)) \times \mathcal{C}_c^\infty(\mathrm{G}_{W^\flat}(F))$ there exists a transfer $f' \in \mathcal{C}_c^\infty(\mathrm{G}'(F))$.*

This was conjectured by Jacquet and Rallis [2011]. For archimedean local fields F'/F, an "approximate transfer" is proved by Xue [n.d.].

5.2.3 Spectral terms: spherical characters. We are now back to F being a number field. For the triple $(\mathrm{G}, \mathrm{H}_1, \mathrm{H}_2)$ we have defined the global (resp. local) spherical characters by (2-8) (resp. by (2-10)). For the triple $(\mathrm{G}', \mathrm{H}_1', \mathrm{H}_2')$, we define the global spherical character associated to a cuspidal automorphic representation Π of $\mathrm{G}'(\mathbb{A})$,

$$\mathbb{J}_\Pi(f', s) = \sum_{\phi \in \mathrm{OB}(\Pi)} \mathcal{P}_{\mathrm{H}_1', s}(\Pi(f)\phi)\overline{\mathcal{P}_{\mathrm{H}_2', \eta}(\phi)}, \quad f' \in \mathcal{C}_c^\infty(\mathrm{G}'(\mathbb{A})), \quad s \in \mathbb{C},$$

where $\mathcal{P}_{\mathrm{H}_1', s}$ is the automorphic period integral $\mathcal{P}_{\mathrm{H}_1', \chi_s}$ for the character χ_s of $\mathrm{H}_1(\mathbb{A}) \in \mathrm{GL}_{n-1}(\mathbb{A}_{F'})$ defined by $h \mapsto |\det(h)|_{F'}^s$. We set $\mathbb{J}_\Pi(f') = \mathbb{J}_\Pi(f', 0)$. We expect to have a global character identity W. Zhang [2014a, Conj. 4.2]:

Conjecture 5.6. *Let π be a tempered cuspidal automorphic representation of $\mathrm{G}(\mathbb{A})$ such that $\mathrm{Hom}_{\mathrm{H}(\mathbb{A})}(\pi, \mathbb{C}) \neq 0$. Let $\Pi = \mathrm{BC}(\pi)$ be its base change which we assume is cuspidal. Then for $f' \in \mathcal{C}_c^\infty(\mathrm{G}'(\mathbb{A}))$ and any transfer $f \in \mathcal{C}_c^\infty(\mathrm{G}(\mathbb{A}))$,*

$$\mathbb{I}_\pi(f) = 2^{-\beta_\pi}\,\mathbb{J}_\Pi(f').$$

Remark 5.7. In fact here we have $\beta_\pi = 2$ due to the cuspidality of $\Pi = \mathrm{BC}(\pi)$. Conjecture 5.6 is known when π_v is supercuspidal at a place v split in F'/F, cf. W. Zhang

[2014b]. In general it should follow from a full spectral decomposition of the Jacquet–Rallis relative trace formulas, and perhaps along the way one will discover the correct definition of the global spherical character $\mathbb{J}_\Pi(f')$ when $\Pi = \mathrm{BC}(\pi)$ is not cuspidal.

In W. Zhang [2014a, §3.4], we defined a local spherical character $\mathbb{J}_{\Pi_v}(f'_v, s)$ for any tempered Π_v (depending on some auxiliary data). Then, for pure tensors $f' = \otimes_v f'_v$, the global spherical character decomposes naturally as an Euler product for $\Pi = \mathrm{BC}(\pi)$,

$$\mathbb{J}_\Pi(f', s) = 2^{-\beta_\pi} \mathscr{L}(s + 1/2, \pi) \prod_v \mathbb{J}_{\Pi_v}(f'_v, s). \tag{5-9}$$

We set $\mathbb{J}_{\Pi_v}(f'_v) = \mathbb{J}_{\Pi_v}(f'_v, 0)$. We expect to have a local character identity W. Zhang [ibid., Conj. 4.4]:

Conjecture 5.8. *Let π_v be a tempered representation of $\mathrm{G}(F_v)$ such that $\mathrm{Hom}_{\mathrm{H}(F_v)}(\pi, \mathbb{C}) \neq 0$. Let $\Pi_v = \mathrm{BC}(\pi_v)$ be its base change. Then for $f'_v \in \mathscr{C}_c^\infty(\mathrm{G}'(F_v))$ and any transfer $f_v \in \mathscr{C}_c^\infty(\mathrm{G}(F_v))$,*

$$\mathbb{J}_{\Pi_v}(f'_v) = \kappa_{\pi_v}\, \mathbb{I}_{\pi_v}(f_v),$$

where κ_{π_v} is an explicit constant.

By a theorem of Harish-Chandra, the character of an admissible representation of $\mathrm{G}(F_v)$ for any reductive group G over a p-adic local field F_v admits a local expansion around the identity of $\mathrm{G}(F_v)$ as a sum of Fourier transforms of unipotent orbital integrals. We have a partial analog for the local spherical character $\mathbb{J}_{\Pi_v}$.

Theorem 5.9. *Let v be a non-archimedean place. Let $\Pi_v = \mathrm{BC}(\pi_v)$ be the base change of a tempered representation π_v of $\mathrm{G}(F_v)$. For any small neighborhood of the identity element of $\mathrm{G}'(F_v)$, there exists an admissible (in the sense of W. Zhang [ibid., §8.1]) function $f'_v \in \mathscr{C}_c^\infty(\mathrm{G}'(F_v))$ such that*

$$\mathbb{J}_{\Pi_v}(f'_v) = c_\Pi \cdot \widehat{\mu}_{\mathrm{reg}}(f'_v),$$

where $\widehat{\mu}_{\mathrm{reg}}$ is the Fourier transform of the (relative) regular unipotent orbital integral, cf. W. Zhang [ibid., §6.3, §8.2], and c_{Π_v} is an explicit constant depending on Π_v.

We have the following.

Theorem 5.10. *Conjecture 5.8 holds if v is split in F'/F, or F_v is a p-adic local field.*

The case of a split v is rather easy W. Zhang [ibid.]. The case of a supercuspidal representation π_v was proved in W. Zhang [ibid.]. For the general p-adic case, the result is proved by Beuzart-Plessis in Beuzart-Plessis [2016] using Theorem 5.9, a local relative trace formula for Lie algebras in W. Zhang [2014b, §4.1], and a group analog in Beuzart-Plessis [2015b].

Remark 5.11. If $\Pi = \mathrm{BC}(\pi)$ is cuspidal, then Conjecture 5.6 and Conjecture 5.8 together imply Conjecture 2.11.

5.3 Arithmetic RTF. In W. Zhang [2012b], the author introduced a relative trace formula approach to the arithmetic Gan–Gross–Prasad conjecture. Let

$$\partial\mathbb{J}(f') = \frac{d}{ds}\Big|_{s=0}\mathbb{J}(f', s),$$

cf. (5-4). Then the idea is that, in analogy to the usual comparison of two RTFs, we hope to compare the height pairing $\mathrm{Int}(f)$ in (3-4) and $\partial\mathbb{J}(f')$ for $f \in \mathcal{H}(\widetilde{\mathrm{G}}, K^{\circ}_{\widetilde{\mathrm{G}}})$ and ay transfer $f' \in \mathcal{C}_c^{\infty}(\mathrm{G}'(\mathbb{A}))$.

Remark 5.12. Here we note that there is no archimedean component in the test function f on the unitary group side. Implicitly we demand that $f' = \otimes_v f'_v \in \mathcal{C}_c^{\infty}(\mathrm{G}'(\mathbb{A}))$, where f'_∞ is a *Gaussian* test function in the sense of Rapoport, Smithling, and W. Zhang [2017b, §7.3] (equivalently, we complete f by tensoring a distinguished archimedean component f_∞ as in W. Zhang [2012b, §3.2, (3.5)]). We also note that, by the isomorphisms (3-2), the orbits on $\widetilde{\mathrm{G}}(F)_{\mathrm{rs}}$ are in natural bijection with those on $\mathrm{G}(F)_{\mathrm{rs}}$, and all geometric terms related to G in Section 5.2.2 transport to $\widetilde{\mathrm{G}}$. We will not repeat the definitions.

To be able to work in a greater generality than the case the height pairing (3-1) is defined, in Rapoport, Smithling, and W. Zhang [2017b], Rapoport, Smithling, and the author turn to the arithmetic intersection theory $(\ ,\ )_{\mathrm{GS}}$ of Arakelov and Gillet–Soulé Gillet and Soulé [1990, §4.2.10] on the arithmetic Chow group $\widehat{\mathrm{Ch}}^*(\mathfrak{X})$ of a regular proper flat scheme (or DM stack) $\mathfrak{X}$ over $\mathrm{Spec}(O_{F'})$. For certain more general levels $K_{\widetilde{\mathrm{G}}} \subset K^{\circ}_{\widetilde{\mathrm{G}}}$, we construct regular integral models $\mathcal{M}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$ of $\mathrm{Sh}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$ (essentially by adding Drinfeld level structures at split primes to the moduli space $\mathcal{M}_{K^{\circ}_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$, cf. Rapoport, Smithling, and W. Zhang [2017b, §4, §5]). We enhance the arithmetic diagonal cycle to an element $\widehat{z}_K$ in $\widehat{\mathrm{Ch}}^{n-1}(\mathcal{M}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}}))$, and we extend the action R of a smaller Hecke algebra $\mathcal{H}^{\mathrm{spl}}(\widetilde{\mathrm{G}}, K_{\widetilde{\mathrm{G}}}) \subset \mathcal{H}(\widetilde{\mathrm{G}}, K_{\widetilde{\mathrm{G}}})$ on the Chow group of $\mathrm{Sh}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$ to an action $\widehat{R}$ on the arithmetic Chow group. We define

$$\mathrm{Int}(f) = \left(\widehat{R}(f)\widehat{z}_K, \widehat{z}_K\right)_{\mathrm{GS}}, \quad f \in \mathcal{H}^{\mathrm{spl}}(\widetilde{\mathrm{G}}, K_{\widetilde{\mathrm{G}}}). \tag{5-10}$$

We can then state an arithmetic intersection conjecture for the arithmetic diagonal cycle on the global integral model $\mathcal{M}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$ Rapoport, Smithling, and W. Zhang [ibid., §8.1, §8.2].

Conjecture 5.13. *Let* $f \in \mathcal{H}^{\mathrm{spl}}(\widetilde{\mathrm{G}}, K_{\widetilde{\mathrm{G}}})$*, and let* $f' \in \mathcal{C}_c^\infty(\mathrm{G}'(\mathbb{A}))$ *be a transfer of* f*. Then*

$$\mathrm{Int}(f) = -\partial\mathbb{J}(f') - \mathbb{J}(f'_{\mathrm{corr}}),$$

where $f'_{\mathrm{corr}} \in \mathcal{C}_c^\infty(\mathrm{G}'(\mathbb{A}))$ *is a correction function. Furthermore, we may choose* f' *such that* $f'_{\mathrm{corr}} = 0$.

Remark 5.14. A deeper understanding of the local spherical character $\mathbb{J}_{\Pi_v}(f'_v, s)$ (or rather, its derivative) in (5-9), together with the spectral decomposition of $\mathbb{J}(f', s)$, should allow us to deduce Conjecture 3.5 from Conjecture 5.13. We hope to return to this point in the future.

The comparison can be localized for a (large) class of test functions f and f'. Let $f = \otimes_v f_v$ be a pure tensor such that there is a place u_0 of F where f_{u_0} has support in the regular semisimple locus $\widetilde{\mathrm{G}}(F_{u_0})_{\mathrm{rs}}$. Then the cycles $\widehat{R}(f)\widehat{z}_K$ and $\widehat{z}_K$ do not meet in the generic fiber $\mathrm{Sh}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}})$. The arithmetic intersection pairing then localizes as a sum over all places w of F' (note that $F = \mathbb{Q}$)

$$\mathrm{Int}(f) = \sum_w \mathrm{Int}_w(f).$$

Here for a non-archimedean place w, the local intersection pairing $\mathrm{Int}_w(f)$ is defined through the Euler–Poincaré characteristic of a derived tensor product on $\mathcal{M}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}}) \otimes_{O_F} O_{F,w}$, cf. Gillet and Soulé [1990, 4.3.8(iv)].

Similarly, let $f' = \otimes_v f'_v$ be a pure tensor such that there is a place u_0 of F where f'_{u_0} has support in the regular semi-simple locus $\mathrm{G}'(F_{u_0})_{\mathrm{rs}}$. Then we have a decomposition

$$\mathbb{J}(f', s) = \sum_{\gamma \in [\mathrm{G}'(F)_{\mathrm{rs}}]} \mathrm{Orb}(\gamma, f', s),$$

where each term is a product of local orbital integrals (5-6),

$$\mathrm{Orb}(\gamma, f', s) = \prod_v \mathrm{Orb}(\gamma, f'_v, s).$$

The first derivative $\partial\mathbb{J}(f')$ then localizes as a sum over places v of F,

$$\partial\mathbb{J}(f') = \sum_v \partial\mathbb{J}_v(f'),$$

where the summand $\partial\mathbb{J}_v(f')$ takes the derivative of the local orbital integral (cf. (5-8)) at the place v,

$$\partial\mathbb{J}_v(f') = \sum_{\gamma \in [\mathrm{G}'(F)_{\mathrm{rs}}]} \partial\mathrm{Orb}(\gamma, f'_v) \cdot \prod_{u \neq v} \mathrm{Orb}(\gamma, f'_u).$$

It is then natural to expect a place-by-place comparison between

$$\mathrm{Int}_v(f) = \sum_{w|v} \mathrm{Int}_w(f) \qquad \text{and} \qquad \partial\mathbb{J}_v(f')$$

If a place v_0 of F splits into two places $w_0, \overline{w}_0$ of F' (and under the above regularity condition on the support of f and of f'), we have Rapoport, Smithling, and W. Zhang [2017b, Thm. 1.3]

$$\mathrm{Int}_{w_0}(f) = \mathrm{Int}_{\overline{w}_0}(f) = \partial\mathbb{J}_{v_0}(f') = 0.$$

For a place w_0 of F' above a non-split place v_0 of $F = \mathbb{Q}$, we have a smooth integral model $\mathfrak{M}_{K_{\widetilde{\mathrm{G}}}}(\widetilde{\mathrm{G}}) \otimes_{O_F} O_{F,w_0}$ when $K_{\widetilde{\mathrm{G}},v_0}$ is a hyperspecial compact open subgroup $\widetilde{\mathrm{G}}(O_{F,v_0})$ (resp. a special parahoric subgroup $K^{\circ}_{\widetilde{\mathrm{G}},v_0}$) for inert v_0 (resp. ramified v_0).

For an inert v_0, the comparison between $\mathrm{Int}_{v_0}(f)$ and $\partial\mathbb{J}_{v_0}(f')$ is then reduced to a local conjecture that we will consider in the next section, the *arithmetic fundamental lemma* conjecture. Let $W^\flat[v_0]$ be the pair of nearby Hermitian spaces, i.e., the Hermitian space (with respect to F'/F) that is totally negative at archimedean places, and is not equivalent to W at v_0. Let $\widetilde{\mathrm{G}}^\flat[v_0]$ be the corresponding group, an inner form of $\widetilde{\mathrm{G}}$.

Theorem 5.15. *Let $f = \otimes_v f_v$ be a pure tensor such that*

(i) $f_{v_0} = \mathbf{1}_{\widetilde{\mathrm{G}}(O_{F,v_0})}$, *and*

(ii) *there is a place u_0 of F where f_{u_0} has support in the regular semisimple locus* $\widetilde{\mathrm{G}}(F_{v_0})_{\mathrm{rs}}$.

Then

$$\mathrm{Int}_{v_0}(f) = \sum_{g \in \widetilde{\mathrm{G}}^\flat[v_0](F)_{\mathrm{rs}}} \mathrm{Int}_{v_0}(g) \cdot \prod_{u \neq v_0} \mathrm{Orb}(g, f_u), \tag{5-11}$$

where the local intersection number $\mathrm{Int}_{v_0}(g)$ *is defined by* (6-2) *in the next section.*

For $F = \mathbb{Q}$, this is W. Zhang [2012b, Thm. 3.9]. The general case is established in (the proof of) Rapoport, Smithling, and W. Zhang [2017b, Thm. 8.15].

The expansion (5-11) resembles the geometric side of a usual RTF, and hence we call the expansion (the geometric side of) an *arithmetic RTF*.

Finally, for a ramified place v_0, the analogous question is reduced to the local *arithmetic transfer* conjecture formulated by Rapoport, Smithling and the author in Rapoport, Smithling, and W. Zhang [2017a] and Rapoport, Smithling, and W. Zhang [2018]. We have a result similar to Theorem 5.15, Rapoport, Smithling, and W. Zhang [2017b, Thm. 8.15].

5.4 Geometric RTF (over function fields). In the last part of this section, let us briefly recall the strategy to prove the higher Gross–Zagier formula in Section 4 over function fields.

To continue from Section 4, let f be an element in the spherical Hecke algebra $\mathscr{H}$ (with $\mathbb{Q}$-coefficient). Let

$$\mathrm{Int}_r(f) := \Big(R(f) * \theta_*^{\mu}[\mathrm{Sht}_T^{\mu}], \quad \theta_*^{\mu}[\mathrm{Sht}_T^{\mu}]\Big)_{\mathrm{Sht}_G'^{\mu}}$$

be the intersection number of the Heegner–Drinfeld cycle with its translation by a Hecke correspondence $R(f)$. Here the right hand side does not depend on μ but only on the number r of modifications of the Shtukas.

Next, consider the triple $(\mathrm{G}', \mathrm{H}_1', \mathrm{H}_2')$ where $\mathrm{G}' = \mathrm{G} = \mathrm{PGL}_2$ and $\mathrm{H}_1' = \mathrm{H}_2'$ are the diagonal torus A of PGL_2. Similarly to 5-4, we define a distribution by a (regularized) integral

$$\mathbb{J}(f,s) = \int_{[\mathrm{H}_1']} \int_{[\mathrm{H}_2']} K_f(h_1,h_2)\, |h_1 h_2|^s\, \eta(h_2)\, dh_1\, dh_2,$$

where, for $h = \mathrm{diag}(a,d) \in A(\mathbb{A})$, we write $|h| = |a/d|$ and $\eta(h) = \eta_{F'/F}(a/d)$. Let

$$\mathbb{J}_r(f) = \frac{d^r}{ds^r}\Big|_{s=0} \mathbb{J}(f,s).$$

Then Yun and the author proved in Yun and W. Zhang [2017] the following *key identity*, which we may call a *geometric RTF*, in contrast to the arithmetic intersection numbers in the number field case.

Theorem 5.16. *Let $f \in \mathscr{H}$. Then*

$$\mathbb{I}_r(f) = (\log q)^{-r} \mathbb{J}_r(f). \tag{5-12}$$

In this situation of geometric intersection, our proof of the key identity (5-12) is entirely global, in the sense that we do *not* reduce the identity to the comparison of local orbital integrals. In fact, our proof of (5-12) gives a term-by-term identity of the orbital integrals. This strategy is explained in the forthcoming work of Yun on the function field analog of the arithmetic fundamental lemma Yun [n.d.(b)].

For a more detailed exposition on the geometric construction related to the proof of Equation (5-12), see Yun's survey Yun [n.d.(a)].

6 The arithmetic fundamental lemma conjecture

We now consider the local version of special cycles on Shimura varieties, i.e., (formal) cycles on Rapoport–Zink formal moduli spaces of p-divisible groups. The theorem of

Rapoport–Zink on the uniformization of Shimura varieties Rapoport and Zink [1996] relates the local cycles to the global ones, and this allows us to relate the local height of the global cycles (the *semi-global* situation in Rapoport, Smithling, and W. Zhang [2017b]) to intersection numbers of local cycles, cf. Theorem 5.15.

6.1 The fundamental lemma of Jacquet and Rallis, and a theorem of Yun. Now let F be a finite extension of $\mathbb{Q}_p$ for an *odd* prime p. Let O_F be the ring of integers in F, and denote by q the number of elements in the residue field of O_F. Let $\breve{F}$ be the completion of a maximal unramified extension of F. Let F'/F be an *unramified* quadratic extension.

Recall from Section 5.2.2 that there are two equivalence classes of pairs of Hermitian spaces, denoted by $W, W^\flat$ respectively, such that, for $W = (W_{n-1}, W_n)$, both W_{n-1} and W_n contain self-dual lattices. We rename the respective groups as G and $\mathrm{G}^\flat$ respectively, and we rewrite the bijection of orbits (5-5):

$$\left[\mathrm{G}(F)_{\mathrm{rs}}\right] \sqcup \left[\mathrm{G}^\flat(F)_{\mathrm{rs}}\right] \xrightarrow{\sim} \left[\mathrm{G}'(F)_{\mathrm{rs}}\right]. \tag{6-1}$$

Let $\mathrm{G}(O_F)$ be the hyperspecial compact open subgroup of $\mathrm{G}(F)$ defined by a self-dual lattice.

Theorem 6.1 (Fundamental Lemma (FL)). *For a prime p sufficiently large, the characteristic function $\mathbf{1}_{\mathrm{G}'(O_{F_0})} \in \mathcal{C}_c^\infty(\mathrm{G}'(F))$ transfers to the pair of functions $(\mathbf{1}_{\mathrm{G}(O_F)}, 0) \in \mathcal{C}_c^\infty(\mathrm{G}(F)) \times \mathcal{C}_c^\infty(\mathrm{G}^\flat(F))$.*

Jacquet and Rallis conjecture that the same is always true for any *odd* p Jacquet and Rallis [2011]. Yun proved the equal characteristic analog of their conjecture for $p > n$; Gordon deduced the p-adic case for p large (but unspecified), cf. Yun [2011].

6.2 The arithmetic fundamental lemma conjecture. Next, we let $\mathcal{N}_n$ be the unitary Rapoport–Zink formal moduli space over $\operatorname{Spf} O_{\breve{F}}$ parameterizing *Hermitian supersingular formal $O_{F'}$-modules of signature* $(1, n-1)$, cf. S. Kudla and Rapoport [2011] and Rapoport, Smithling, and W. Zhang [2018]. Let $\mathcal{N}_{n-1,n} = \mathcal{N}_{n-1} \times_{\operatorname{Spf} O_{\breve{F}}} \mathcal{N}_n$. Then $\mathcal{N}_{n-1,n}$ admits an action by $\mathrm{G}^\flat(F)$.

There is a natural closed embedding $\delta \colon \mathcal{N}_{n-1} \to \mathcal{N}_n$ (a local analog of the closed embedding (3-3)). Let

$$\Delta \colon \mathcal{N}_{n-1} \longrightarrow \mathcal{N}_{n-1,n}$$

be the graph morphism of δ. We denote by $\Delta_{\mathcal{N}_{n-1}}$ the image of Δ. For $g \in \mathrm{G}^\flat(F)_{\mathrm{rs}}$, we consider the intersection product on $\mathcal{N}_{n-1,n}$ of $\Delta_{\mathcal{N}_{n-1}}$ with its translate $g\Delta_{\mathcal{N}_{n-1}}$, defined through the derived tensor product of the structure sheaves,

$$\operatorname{Int}(g) := (\Delta_{\mathcal{N}_{n-1}}, g \cdot \Delta_{\mathcal{N}_{n-1}})_{\mathcal{N}_{n-1,n}} := \chi\left(\mathcal{N}_{n-1,n}, \mathcal{O}_{\Delta_{\mathcal{N}_{n-1}}} \otimes^{\mathbb{L}} \mathcal{O}_{g\cdot\Delta_{\mathcal{N}_{n-1}}}\right). \tag{6-2}$$

We have defined the derivative of the orbital integral (5-8).

Conjecture 6.2 (Arithmetic Fundamental Lemma (AFL), W. Zhang [2012b]). *Let* $\gamma \in \mathrm{G}'(F)_{\mathrm{rs}}$ *match an element* $g \in \mathrm{G}^{\flat}(F)_{\mathrm{rs}}$. *Then*

$$\omega(\gamma)\, \partial\mathrm{Orb}\left(\gamma, \mathbf{1}_{\mathrm{G}'(O_F)}\right) = -2\,\mathrm{Int}(g) \cdot \log q.$$

Remark 6.3. (i) We may interpret the orbital integrals in terms of "counting lattices", cf. Rapoport, Terstiege, and W. Zhang [2013, §7].

(ii) See Rapoport, Smithling, and W. Zhang [2018, §4] for some other equivalent formulations of the AFL conjecture.

Theorem 6.4. (i) *The AFL Conjecture 6.2 holds when* $n = 2, 3$.

(ii) *When* $p \geq \frac{n}{2} + 1$, *the AFL Conjecture 6.2 holds for* minuscule *elements* $g \in \mathrm{G}^{\flat}(F)$ *in the sense of Rapoport, Terstiege, and W. Zhang [2013].*

Part (i) was proved in W. Zhang [2012b]; a simplified proof when $p \geq 5$ is given by Mihatsch in Mihatsch [2017]. Part (ii) was proved by Rapoport, Terstiege, and the author in Rapoport, Terstiege, and W. Zhang [2013]; a simplified proof is given by Li and Zhu in Li and Y. Zhu [2017]. Mihatsch in Mihatsch [2016] proved more cases of the AFL for arbitrary n but under restrictive conditions on g.

Remark 6.5. (i) Yun has announced a proof of the function field analog of the AFL conjecture Yun [n.d.(a),(b)].

(ii) Let F'/F be a *ramified* quadratic extension of p-adic fields. In Rapoport, Smithling, and W. Zhang [2017a] and Rapoport, Smithling, and W. Zhang [2018], Rapoport, Smithling, and the author propose an *arithmetic transfer* (AT) conjecture. This conjecture can be viewed as the counterpart of the existence of transfer (cf. Theorem 5.5) in the arithmetic context over a p-adic field. We proved the conjecture for $n = 2, 3$.

(iii) The analogous question on archimedean local fields remains a challenge, involving Green currents in the complex geometric setting and relative orbital integrals on real Lie groups.

References

Avraham Aizenbud, Dmitry Gourevitch, Stephen Rallis, and Gérard Schiffmann (2010). "Multiplicity one theorems". *Ann. of Math. (2)* 172.2, pp. 1407–1434. MR: 2680495 (cit. on p. 512).

R. Beuzart-Plessis (2015a). "Endoscopie et conjecture locale raffinée de Gan-Gross-Prasad pour les groupes unitaires". *Compos. Math.* 151.7, pp. 1309–1371. MR: 3371496 (cit. on p. 513).

Raphaël Beuzart-Plessis (June 2015b). "A local trace formula for the Gan-Gross-Prasad conjecture for unitary groups: the archimedean case". arXiv: 1506.01452 (cit. on pp. 513, 532).

– (Feb. 2016). "Comparison of local spherical characters and the Ichino-Ikeda conjecture for unitary groups". arXiv: 1602.06538 (cit. on pp. 514, 532).

Pierre-Henri Chaudouard and Michał Zydor (Nov. 2016). "Le transfert singulier pour la formule des traces de Jacquet-Rallis". arXiv: 1611.09656 (cit. on p. 513).

Pierre Deligne (1971). "Travaux de Shimura", 123–165. Lecture Notes in Math., Vol. 244. MR: 0498581 (cit. on p. 516).

Solomon Friedberg and Hervé Jacquet (1993). "Linear periods". *J. Reine Angew. Math.* 443, pp. 91–139. MR: 1241129 (cit. on p. 511).

W. T. Gan, B. H. Gross, D. Prasad, and J.-L. Waldspurger (2012). *Sur les conjectures de Gross et Prasad. I*. Astérisque No. 346 (2012) (2012). Société Mathématique de France, Paris, pp. xi+318. MR: 3052279 (cit. on p. 513).

Wee Teck Gan (2014). "Recent progress on the Gross-Prasad conjecture". *Acta Math. Vietnam.* 39.1, pp. 11–33. MR: 3176460 (cit. on p. 514).

Wee Teck Gan, Benedict H. Gross, and Dipendra Prasad (2012). "Symplectic local root numbers, central critical L values, and restriction problems in the representation theory of classical groups". *Astérisque* 346. Sur les conjectures de Gross et Prasad. I, pp. 1–109. MR: 3202556 (cit. on pp. 507, 512, 513, 518, 520, 529).

Henri Gillet and Christophe Soulé (1990). "Arithmetic intersection theory". *Inst. Hautes Études Sci. Publ. Math.* 72, 93–174 (1991). MR: 1087394 (cit. on pp. 519, 533, 534).

David Ginzburg, Dihua Jiang, and Stephen Rallis (2004). "On the nonvanishing of the central value of the Rankin-Selberg L-functions". *J. Amer. Math. Soc.* 17.3, pp. 679–722. MR: 2053953 (cit. on p. 513).

– (2009). "Models for certain residual representations of unitary groups". In: *Automorphic forms and L-functions I. Global aspects*. Vol. 488. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 125–146. MR: 2522030 (cit. on p. 513).

Benedict H. Gross (1997). "On the motive of a reductive group". *Invent. Math.* 130.2, pp. 287–313. MR: 1474159 (cit. on p. 513).

Benedict H. Gross and Dipendra Prasad (1992). "On the decomposition of a representation of SO_n when restricted to SO_{n-1}". *Canad. J. Math.* 44.5, pp. 974–1002. MR: 1186476 (cit. on p. 512).

– (1994). "On irreducible representations of $\mathrm{SO}_{2n+1} \times \mathrm{SO}_{2m}$". *Canad. J. Math.* 46.5, pp. 930–950. MR: 1295124 (cit. on p. 512).

Benedict H. Gross and Don B. Zagier (1986). "Heegner points and derivatives of L-series". *Invent. Math.* 84.2, pp. 225–320. MR: 833192 (cit. on pp. 507, 517, 524).

G. Harder, R. P. Langlands, and M. Rapoport (1986). "Algebraische Zyklen auf Hilbert-Blumenthal-Flächen". *J. Reine Angew. Math.* 366, pp. 53–120. MR: 833013 (cit. on p. 517).

R. Neal Harris (2014). "The refined Gross-Prasad conjecture for unitary groups". *Int. Math. Res. Not. IMRN* 2, pp. 303–389. MR: 3159075 (cit. on pp. 513, 514).

Hongyu He (2017). "On the Gan-Gross-Prasad conjecture for $U(p, q)$". *Invent. Math.* 209.3, pp. 837–884. MR: 3681395 (cit. on p. 513).

Atsushi Ichino (2008). "Trilinear forms and the central values of triple product L-functions". *Duke Math. J.* 145.2, pp. 281–307. MR: 2449948 (cit. on p. 514).

Atsushi Ichino and Tamutsu Ikeda (2010). "On the periods of automorphic forms on special orthogonal groups and the Gross-Prasad conjecture". *Geom. Funct. Anal.* 19.5, pp. 1378–1425. MR: 2585578 (cit. on pp. 513, 514).

Hervé Jacquet (2005). "A guide to the relative trace formula". In: *Automorphic representations, L-functions and applications: progress and prospects*. Vol. 11. Ohio State Univ. Math. Res. Inst. Publ. de Gruyter, Berlin, pp. 257–272. MR: 2192826 (cit. on pp. 508, 510, 527).

Hervé Jacquet and Stephen Rallis (2011). "On the Gross-Prasad conjecture for unitary groups". In: *On certain L-functions*. Vol. 13. Clay Math. Proc. Amer. Math. Soc., Providence, RI, pp. 205–264. MR: 2767518 (cit. on pp. 529, 531, 537).

Dihua Jiang and Lei Zhang (Aug. 2015). "Arthur Parameters and Cuspidal Automorphic Modules of Classical Groups". arXiv: 1508.03205 (cit. on p. 513).

Stephen S. Kudla (2002). "Derivatives of Eisenstein series and arithmetic geometry". In: *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*. Higher Ed. Press, Beijing, pp. 173–183. MR: 1957031.

Stephen S. Kudla, Michael Rapoport, and Tonghai Yang (2006). *Modular forms and special cycles on Shimura curves*. Vol. 161. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. x+373. MR: 2220359 (cit. on p. 517).

Stephen Kudla and Michael Rapoport (2011). "Special cycles on unitary Shimura varieties I. Unramified local theory". *Invent. Math.* 184.3, pp. 629–682. MR: 2800697 (cit. on p. 537).

– (2014). "Special cycles on unitary Shimura varieties II: Global theory". *J. Reine Angew. Math.* 697, pp. 91–157. MR: 3281653 (cit. on pp. 517, 521).

Klaus Künnemann (2001). "Height pairings for algebraic cycles on abelian varieties". *Ann. Sci. École Norm. Sup. (4)* 34.4, pp. 503–523. MR: 1852008 (cit. on p. 519).

Erez M Lapid (2006). "The relative trace formula and its applications (Automorphic Forms and Automorphic L-Functions)". *Surikaisekikenkyusho Kokyuroku* 1468, pp. 76–87 (cit. on p. 527).

– (2010). “Some applications of the trace formula and the relative trace formula”. In: *Proceedings of the International Congress of Mathematicians*. Vol. III. Hindustan Book Agency, New Delhi, pp. 1262–1280 (cit. on p. 527).

Chao Li and Yihang Zhu (2017). “Remarks on the arithmetic fundamental lemma”. *Algebra Number Theory* 11.10, pp. 2425–2445. MR: 3744362 (cit. on p. 538).

Yifeng Liu (2014). “Relative trace formulae toward Bessel and Fourier-Jacobi periods on unitary groups”. *Manuscripta Math.* 145.1-2, pp. 1–69. MR: 3244725 (cit. on p. 529).

– (2016). “Refined global Gan-Gross-Prasad conjecture for Bessel periods”. *J. Reine Angew. Math.* 717, pp. 133–194. MR: 3530537 (cit. on p. 529).

Andreas Mihatsch (Nov. 2016). “Relative unitary RZ-spaces and the Arithmetic Fundamental Lemma”. arXiv: 1611.06520 (cit. on p. 538).

– (2017). “On the arithmetic fundamental lemma through Lie algebras”. *Math. Z.* 287.1-2, pp. 181–197. MR: 3694673 (cit. on p. 538).

Colette Mœglin and Jean-Loup Waldspurger (2012). “La conjecture locale de Gross-Prasad pour les groupes spéciaux orthogonaux: le cas général”. *Astérisque* 347. Sur les conjectures de Gross et Prasad. II, pp. 167–216. MR: 3155346 (cit. on p. 513).

Omer Offen (2009). “Unitary periods and Jacquet’s relative trace formula”. In: *Automorphic forms and L-functions I. Global aspects*. Vol. 488. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 183–236. MR: 2522031 (cit. on pp. 511, 527).

M. Rapoport, B. Smithling, and W. Zhang (2017a). “On the arithmetic transfer conjecture for exotic smooth formal moduli spaces”. *Duke Math. J.* 166.12, pp. 2183–2336. MR: 3694568 (cit. on pp. 508, 530, 535, 538).

M. Rapoport and Th. Zink (1996). *Period spaces for p-divisible groups*. Vol. 141. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xxii+324. MR: 1393439 (cit. on p. 537).

Michael Rapoport, Brian Smithling, and Wei Zhang (Oct. 2017b). “Arithmetic diagonal cycles on unitary Shimura varieties”. arXiv: 1710.06962 (cit. on pp. 519–522, 533, 535, 537).

– (2018). “Regular formal moduli spaces and arithmetic transfer conjectures”. *Math. Ann.* 370 (cit. on pp. 508, 531, 535, 537, 538).

Michael Rapoport, Ulrich Terstiege, and Wei Zhang (2013). “On the arithmetic fundamental lemma in the minuscule case”. *Compos. Math.* 149.10, pp. 1631–1666. MR: 3123304 (cit. on p. 538).

Michael Rapoport and Eva Viehmann (2014). “Towards a theory of local Shimura varieties”. *Münster J. Math.* 7.1, pp. 273–326. MR: 3271247 (cit. on p. 516).

Yiannis Sakellaridis (2008). “On the unramified spectrum of spherical varieties over p-adic fields”. *Compos. Math.* 144.4, pp. 978–1016. MR: 2441254 (cit. on p. 510).

Yiannis Sakellaridis and Akshay Venkatesh (Feb. 2012). “Periods and harmonic analysis on spherical varieties”. To appear in Astérisque, 296pp. arXiv: 1203.0039 (cit. on pp. 510, 514).
Binyong Sun and Chen-Bo Zhu (2012). “Multiplicity one theorems: the Archimedean case”. *Ann. of Math. (2)* 175.1, pp. 23–44. MR: 2874638 (cit. on p. 512).
J.-L. Waldspurger (1985). “Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie”. *Compositio Math.* 54.2, pp. 173–242. MR: 783511 (cit. on pp. 507, 511).
H. Xue (n.d.). “On the global Gan–Gross–Prasad conjecture for unitary groups: approximating smooth transfer of Jacquet–Rallis”. To appear in *J. Reine Angew. Math.* (cit. on pp. 513, 531).
Xinyi Yuan, Shou-Wu Zhang, and Wei Zhang (n.d.). “Triple product L-series and Gross-Kudla-Schoen cycles”. In preparation (cit. on p. 520).
– (2013). *The Gross-Zagier formula on Shimura curves*. Vol. 184. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. x+256. MR: 3237437 (cit. on pp. 517, 520, 524).
Zhiwei Yun (n.d.[a]). “Hitchin type moduli stacks in automorphic representation theory”. ICM report. (cit. on pp. 536, 538).
– (n.d.[b]). “The arithmetic fundamental lemma over function fields”. In preparation (cit. on pp. 536, 538).
– (2011). “The fundamental lemma of Jacquet and Rallis”. *Duke Math. J.* 156.2. With an appendix by Julia Gordon, pp. 167–227. MR: 2769216 (cit. on p. 537).
Zhiwei Yun and Wei Zhang (n.d.). “Shtukas and the Taylor expansion of L-functions (II)”. Preprint (cit. on pp. 526, 527).
– (2017). “Shtukas and the Taylor expansion of L-functions”. *Ann. of Math. (2)* 186.3, pp. 767–911. MR: 3702678 (cit. on pp. 508, 523, 525, 526, 536).
Shouwu Zhang (2010). “Linear forms, algebraic cycles, and derivatives of L-series”. preprint (cit. on p. 522).
Wei Zhang (2009). “Relative trace formula and arithmetic Gross-Prasad conjecture”. Unpublished manuscript (cit. on p. 522).
– (2012a). “Gross-Zagier formula and arithmetic fundamental lemma”. In: *Fifth International Congress of Chinese Mathematicians. Part 1, 2*. Vol. 2. AMS/IP Stud. Adv. Math., 51, pt. 1. Amer. Math. Soc., Providence, RI, pp. 447–459. MR: 2908086.
– (2012b). “On arithmetic fundamental lemmas”. *Invent. Math.* 188.1, pp. 197–252. MR: 2897697 (cit. on pp. 508, 529, 530, 533, 535, 538).
– (2014a). “Automorphic period and the central value of Rankin-Selberg L-function”. *J. Amer. Math. Soc.* 27.2, pp. 541–612. MR: 3164988 (cit. on pp. 511, 514–516, 531, 532).

– (2014b). "Fourier transform and the global Gan-Gross-Prasad conjecture for unitary groups". *Ann. of Math. (2)* 180.3, pp. 971–1049. MR: 3245011 (cit. on pp. 513, 531, 532).

Michał Zydor (Oct. 2015). "Les formules des traces relatives de Jacquet-Rallis grossières". arXiv: 1510.04301 (cit. on p. 513).

Received 2017-12-14.

Wei Zhang (张伟)
Massachusetts Institute of Technology
Department of Mathematics
77 Massachusetts Avenue
Cambridge, MA 02139
USA
weizhang@mit.edu

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (541–564)

RESOLUTION OF SINGULARITIES OF COMPLEX ALGEBRAIC VARIETIES AND THEIR FAMILIES

DAN ABRAMOVICH

Abstract

We discuss Hironaka's theorem on resolution of singularities in charactetistic 0 as well as more recent progress, both on simplifying and improving Hironaka's method of proof and on new results and directions on families of varieties, leading to joint work on toroidal orbifolds with Michael Temkin and Jarosław Włodarczyk.

Contents

1 Introduction

1.1 Varieties and singularities. An affine complex algebraic variety X is the zero set in $\mathbb{C}^n$ of a collection of polynomials $f_i \in \mathbb{C}[x_1, \ldots, x_n]$, and a general complex algebraic

New research reported here is supported by BSF grant 2014365, which is held jointly with J. Włodarczyk (Purdue) and M. Temkin (Jerusalem).

MSC2010: primary 14E15; secondary 14M25, 14A20.

variety is patched together from such affine varieties much as a differentiable manifold is patched together from euclidean balls.

But unlike differentiable manifolds, which locally are all the same (given the dimension), a complex algebraic variety can have an interesting structure locally at a point $p \in X$: the point is *regular* or *simple* if the f_i form the defining equations of a differentiable submanifold, and otherwise it is *singular*, hiding a whole world within it. In the case of one equation, a point $p = (a_1, \ldots, a_n)$ is regular precisely when the defining equation f_1 has a non-vanishing derivative at p, and in general one needs to look at the Jacobian matrix of the defining equations, just like when studying submanifolds. The set of regular points $X^{reg} \subset X$ is always open. The variety X is itself *regular* if $X^{reg} = X$.[1]

1.2 What is resolution of singularities? A look at a few singularities[2] quickly reveals that they are quite beautiful, but complicated - really they are not *simple*. How can we understand them? *Resolution of singularities* provides one approach. For simplicity we restrict to *irreducible* varieties, namely those which cannot be written as a union $X_1 \cup X_2$ of two closed nonempty subvarieties. A resolution of singularities of a variety X is a surgery operation, a morphism $X' \to X$ which takes out the singular points and replaces them by regular points. Formally:

Definition 1.2.1. *A* resolution of singularities *of an irreducible variety X is a* proper morphism $f : X' \to X$, *where X' is regular and irreducible, and f restricts to an isomorphism $f^{-1}(X^{reg}) \longrightarrow X^{reg}$.*

The irreducibility assumption is not serious — for instance one can resolve each irreducible component separately.

I need to explain the terms. A *morphism* $f : X' \to X$ is a mapping which locally on affine patches is defined by polynomials of the coordinates. It is an isomorphism if it is invertible as such. It is *proper* if it is proper as a mapping of topological spaces in the usual Euclidean topology: the image of a compact subset is compact. This is a way to say that we are missing no points: it would be cheating - and useless - to define X' to be just X^{reg}: the idea is to *parametrize* X in a way that reveals the depths of its singularities - not to erase them! One way $X' \to X$ can be guaranteed to be proper is if it is *projective*, namely X' embeds as a closed subvariety of $X \times \mathbb{P}^n$ for some n, where $\mathbb{P}^n$ stands for the complex projective space.

1.3 Hironaka's theorem. In 1964 Hironaka published the following, see Hironaka [1964, Main Theorem 1]:

[1] I am taking the scheme theoretic approach here: the vanishing locus of $f(x) = x^2$ is singular.

[2] I cannot improve on these:
https://imaginary.org/gallery/herwig-hauser-classic

Theorem 1.3.1 (Hironaka). *Let X be a complex algebraic variety. Then there is a projective resolution of singularities $X' \to X$.*

Hironaka's theorem is an end of an era, but also a beginning: in the half century since, people, including Hironaka, have continued to work with renewed vigor on resolution of singularities. Why is that?

I see two reasons. One reason can be seen in Grothendieck's address, Grothendieck [1971]:

> Du point de vue technique, la démonstration du théorème de Hironaka constitue une prouesse peu commune. Le rapporteur avoue n'en avoir pas fait entièrement le tour. Aboutissement d'années d'efforts concentrés, elle est sans doute l'une des démonstrations les plus «dures» et les plus monumentales qu'on connaisse en mathématique.

Consider, for instance, that Hironaka developed much of the theory now known as *Gröbner bases* (at roughly the same time as Buchberger [1965]) for the purpose of resolution of singularities!

There has been a monumental effort indeed to simplify Hironaka's proof, and to break it down to more basic elements, so that the techniques involved come naturally and the ideas flow without undue effort. I think this has been a resounding success and Grothendieck himself would have approved of the current versions of the proof - he certainly would no longer have trouble going through it. In my exposition I attempt to broadly describe the results of this effort.

A few points in this effort are marked by the following:

- The theory of maximal contact: Giraud [1974] and Aroca, Hironaka, and Vicente [1977].
- Constructive resolution using an invariant: Villamayor [1989], Bierstone and Milman [1997], and Encinas and Villamayor [1998].
- The optimal version of canonical resolution: Bierstone and Milman [1997].
- Simplification using order of ideal: Encinas and Villamayor [2003].
- Functoriality as a proof technique and guiding principle: Włodarczyk [2005] and Bierstone and Milman [2008].
- Dissemination to the masses: Cutkosky [2004], Hauser [2006], and Kollár [2007].

The other reason is generalizations and refinements of the resolution theorem. First and foremost, algebraic geometers want to resolve singularities in positive and mixed characteristics, as the implications would be immense. In addition, one is interested in simplifying families of varieties, simplifying algebraic differential equations, making the resolution process as effective and as canonical as possible, and preserving structure one is provided with at the outset. Below I will discuss *resolution in families* while preserving *toroidal* structures, focussing on joint work with Michael Temkin of Jerusalem and Jarosław Włodarczyk of Purdue.

2 Hironaka's method: from resolution to order reduction

The purpose of this section is to indicate how Hironaka's resolution of singularities can be reduced to an algebraic problem, namely *order reduction* of an ideal.

2.1 Blowing up. The key tool for Hironaka's resolution of singularities is an operation called *blowing up* of a regular subvariety Z of a regular variety Y, see Hartshorne [1977, Definition p. 163]

2.1.1 Blowing up a point. A good idea can be gleaned from the special case where $Y = \mathbb{A}^n$, affine n-space, and Z is the origin, as explained in Hartshorne [ibid., Example 7.12.1] and depicted on the cover of Shafarevich [2013]. Think about $\mathbb{P}^{n-1}$ as the set of lines in $\mathbb{A}^n$ through the origin. The blowing up $Y' \to Y$ is then given as the *incidence variety*

$$Y' = \{(x, \ell) \in \mathbb{A}^n \times \mathbb{P}^{n-1} \mid x \in \ell\},$$

with its natural projection to $\mathbb{A}^n$.

This can be described in equations as follows:

$$Y' = \{((x_1, \dots, x_n), (Y_1 : \dots : Y_n)) \in \mathbb{A}^n \times \mathbb{P}^{n-1} \mid x_i Y_j = x_j Y_i \ \forall i, j\}.$$

Since $\mathbb{P}^n$ is covered by affine charts, this can further be simplified. For instance on the chart where $Y_n \neq 0$ with coordinates $y_1, \dots, y_{n-1}$, this translates to

$$\{((x_1, \dots, x_n), (y_1, \dots, y_{n-1})) \in \mathbb{A}^n \times \mathbb{A}^{n-1} \mid x_j = x_n y_j,\ 1 \leq j \leq n-1\}.$$

In other words, the coordinates $x_1, \dots, x_{n-1}$ are redundant and this is just affine space with coordinates $y_1, \dots, y_{n-1}, x_n$. In terms of coordinates on Y we have $y_j = x_j / x_n$.

The fibers of $Y' \to Y$ are easy to describe: away from the origin $x_1 = \dots = x_n = 0$ the map is invertible, as the line ℓ is uniquely determined by $(x_1, \dots, x_n)$. Over the origin all possible lines occur, so the fiber is $\mathbb{P}^{n-1}$, naturally identified as the space of lines through the origin.

2.1.2 Blowing up a regular subvariety. In general the process is similar: given regular subvariety Z of Y, then $f : Y' \to Y$ replaces each point $z \in Z$ by the projective space of normal directions to Z at z. If Z is locally defined by equations $x_1 = \ldots = x_k = 0$ and if $x_{k+1}, \ldots, x_n$ form coordinates along Z, then Y' has local patches, one corresponding to each x_i, with coordinates

$$\frac{x_1}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \; x_i, \; \frac{x_{i+1}}{x_i}, \ldots, \frac{x_k}{x_i}, \; x_{k+1}, \ldots, x_n.$$

Thus the blowing up $Y' \to Y$ of a regular subvariety Z of a regular variety Y always results in a regular variety.

We often say that Z is the *center* of the blowing up $Y' \to Y$, or that the blowing up $Y' \to Y$ is *centered* at Z.

2.1.3 The Proj **construction.** Grothendieck gave a more conceptual construction, which applies to an arbitrary subscheme Z defined by an ideal sheaf $\mathcal{I}$ in an arbitrary scheme Y:

$$Y' = \operatorname{Proj}_Y \bigoplus_{k=0}^{\infty} \mathcal{I}^k / \mathcal{I}^{k+1}.$$

The map $f^{-1}(Y \setminus Z) \to (Y \setminus Z)$ is always an isomorphism, so we identify $Y \setminus Z$ with its preimage.

What the reader may want to take from this is that the blowings up we introduced in particularly nice cases are part of a flexible array of transformations.

The complement E of $Y \setminus Z$ in Y' is called the *exceptional locus*. It is *a Cartier divisor*, a subvariety of codimension 1 locally defined by one equation. If Z is nowhere dense in Y, then $Y' \to Y$ is birational. If moreover Y and Z are regular, then E is regular.

2.1.4 The strict transform. Blowing up serves in the resolution of singularities of a subvariety $X \subset Y$ through the *strict transform* $X' \subset Y'$: this is the closure of $X \setminus Z$ in Y'. Grothendieck showed that X' is the same as the blowing up of $X \cap Z$ in X, using the Proj construction above.

From the point of view of resolution of singularities, the challenge is to make X' less singular than X by an appropriate choice of Z.

Consider for instance the cuspidal plane curve X given by $y^2 - x^3 = 0$ in the affine plane $Y = \mathbb{A}^2$ with coordinates x, y. Blowing up the origin and focusing on the chart with coordinates $x, z = y/x$, we obtain the equation $z^2x^2 - x^3 = 0$, which we rewrite as $x^2(z^2 - x) = 0$. The locus $x = 0$ describes the exceptional line, and X' is given by $z^2 - x = 0$, a regular curve.

2.2 Embedded resolution. Theorem 1.3.1 is proven by way of the following theorem:

Theorem 2.2.1 (Embedded resolution)**.** *Suppose $X \subset Y$ is a closed subvariety of a regular variety Y. There is a sequence of blowings up $Y_n \to Y_{n-1} \to \cdots \to Y_0 = Y$, with regular centers $Z_i \subset Y_i$ and strict transforms $X_i \subset Y_i$, such that Z_i does not contain any irreducible component of X_i and such that X_n is regular.*

In other words, the final strict transform of X in a suitably chosen sequence of blowings up of Y is regular.

In the example of the cuspidal curve $X \subset Y = \mathbb{A}^2$, the single blowing up $Y_1 \to Y$ centered at the origin $Z = \{(0,0)\}$ provides an embedded resolution $X' \to X$ of X.

If X is embedded inside a regular variety Y, then Theorem 2.2.1 immediately gives a resolution of singularities $X_n \to X$. What if X is not embedded? There are a number of viable approaches, but the best is to strengthen Theorem 2.2.1: one makes the blowing up procedure independent of re-embedding X and compatible with local patching. This is what is done in practice. I describe the underlying principles in Sections 3.8 and 3.10 below. The upshot is that "good embedded resolution implies resolution".

From here on we pursue a good embedded resolution.

2.3 Normal crossings. To go further one needs to describe a desirable property of the exceptional divisor E_i and its interaction with the center Z_i.

Definition 2.3.1. *We say that a closed subset $E \subset Y$ of a regular variety Y is a* simple normal crossings divisor *if in its decomposition $E = \cup E_j$ into irreducible components, each component E_j is regular, and these components intersect transversally: locally at a point $p \in E$ there are local parameters $x_1, \ldots, x_m$ such that E is the zero locus of a reduced monomial $x_1 \cdots x_k$.*

We further say that E and a regular subvariety Z have normal crossings *if such coordinates can be chosen so that $Z = V(x_{j_1}, \ldots, x_{j_l})$ is the zero set of a subset of these coordinates.*

When the set of coordinates $x_{j_1}, \ldots, x_{j_l}$ is disjoint from $x_1, \ldots, x_k$ the strata of E meet Z transversely, but the definition above allows quite a bit more flexibility.

This definition works well with blowing up: If E is a simple normal crossings divisor, E and Z have normal crossings, $f : Y' \to Y$ is the blowing up of the regular center Z with exceptional divisor E_Z, and $E' = f^{-1}E \cup E_Z$ then E' is a simple normal crossings divisor.

2.4 Principalization. Embedded resolution is proven by way of the following algebraic result:

Theorem 2.4.1 (Principalization)**.** *Let Y be a regular variety and $\mathcal{I}$ an ideal sheaf. There is a sequence of blowings up $Y_n \to Y_{n-1} \to \cdots \to Y_0 = Y$, regular subvarieties $Z_i \subset Y_i, i = 0, \ldots, n-1$ and simple normal crossings divisors $E_i \subset Y_i, i = 1, \ldots, n$ such that*

- *$f_i : Y_{i+1} \to Y_i$ is the blowing up of Z_i for $i = 0, \ldots, n-1$,*
- *E_i and Z_i have normal crossings for $i = 1, \ldots, n-1$,*
- *$\mathcal{I}\mathcal{O}_{Y_i}$ vanishes on Z_i for $i = 0, \ldots, n-1$,*
- *E_{i+1} is the union of $f_i^{-1} E_i$ with the exceptional locus of f_i for $i = 0, \ldots, n-1$*

and such that the resulting ideal sheaf $\mathcal{I}_n = \mathcal{I}\mathcal{O}_{Y_n}$ is an invertible ideal with zero set $V(\mathcal{I}_n)$ supported in E_n.

In local coordinates $x_1, \ldots, x_m$ on Y_n as above, this means that $\mathcal{I}_n = (x_1^{a_1} \cdots x_m^{a_k})$ is locally principal and monomial, hence the name “principalization”. The condition that Z_i have normal crossings with E_i guarantees that E_{i+1} is a simple normal crossings divisor.

2.4.2 Principalization implies embedded resolution. Quoting Kollár [2007, p. 137], principalization implies embedded resolution seemingly “by accident”: suppose for simplicity that X is irreducible, and let the ideal of $X \subset Y$ be $\mathcal{I}$. Since $\mathcal{I}_n$ is the ideal of a divisor supported in the exceptional locus, at some point in the sequence the center Z_i must contain the strict transform X_i of X. Since $\mathcal{I}$ vanishes on Z_i, it follows that Z_i coincides with X_i at least near X_i. In particular X_i is regular!

2.4.3 Are we working too hard? Principalization seems to require “too much” for resolution: why should we care about exceptional divisors which lie outside X? Are we trying too hard?

In the example of the cuspidal curve $X \subset Y = \mathbb{A}^2$ above, the single blowing up $Y_1 \to Y$ at the origin does *not* suffice for principalization: the resulting equation $x^2(z^2 - x) = 0$ with exceptional $\{x = 0\}$ is *not* monomial. One needs no less that *three* more blowings up! I’ll describe just one key affine patch of each:

- Blowing up $x = z = 0$ one gets, in one affine patch where $x = zw$, the equation $z^3w^2(z - w) = 0$, strict transform $X_2 = \{z = w\}$ and exceptional $\{wz = 0\}$.
- Blowing up $z = w = 0$ one gets, in one affine patch where $z = wv$, the equation $v^3w^6(v - 1) = 0$. The exceptional in this patch is $\{wv = 0\}$, with one component (the old $\{z = 0\}$) appearing only in the other patch. The strict transform is $X_3 = \{v = 1\}$.

- In the open set $\{v \neq 0\}$ we blow up $\{v = 1\}$. This actually does nothing, except turning the function $u = v - 1$ into a monomial along $v = 1$, so the equation $v^3 w^6 (v-1) = 0$ at these points can be written as $(u+1)^3 w^6 u = 0$, which in this patch is equivalent to $w^6 u = 0$, a monomial in the exceptional parameters u, w.

The fact that we could blow up $\{v = 1\}$ means that X_3 is regular, giving rather late evidence that we obtained resolution of singularities for X. These "redundant" steps add to the sense that this method works "by accident". It turns out that principalization itself is quite useful in the study of singularities. Also the fact that it provides the prize of resolution is seen as sufficient justification. The discussion in Section 6 will put it in the natural general framework of toroidal structures.

Accident or not, we will continue to pursue principalization.

2.5 Order reduction. Finally, principalization of an ideal is proven by way of *order reduction*.

The *order* $\mathrm{ord}_p(\mathcal{I})$ of an ideal $\mathcal{I}$ at a point p of a regular variety Y is the maximum integer d such that $\mathfrak{m}_p^d \supseteq \mathcal{I}$; here $\mathfrak{m}_p$ is the maximal ideal of p. It tells us "how many times every element of $\mathcal{I}_p$ vanishes at p."

In particular we have $\mathrm{ord}_p(\mathcal{I}) \geq 1$ precisely if $\mathcal{I}$ vanishes at p.

We write $\mathrm{maxord}(\mathcal{I}) = \max\{\mathrm{ord}_p(\mathcal{I}) | p \in X\}$. For instance we have $\mathrm{maxord}(\mathcal{I}) = 0$ if and only if $\mathcal{I}$ is the unit ideal, which vanishes nowhere. Another exceptional case is $\mathrm{maxord}(\mathcal{I}) = \infty$ which happens if $\mathcal{I}$ vanishes on a whole component of Y. We'll ignore that case for now.

Given an integer a, we write $V(\mathcal{I}, a)$ for the locus of points p where $\mathrm{ord}_p(\mathcal{I}) \geq a$. A regular closed subvariety $Z \subset Y$ is said to be $(\mathcal{I}, a)$-*admissible* if and only if $Z \subset V(\mathcal{I}, a)$, in other words, the order of $\mathcal{I}$ at every point of Z is at least a. Admissibility is related to blowings up: if $\mathrm{maxord}(\mathcal{I}) = a$, and if $Y' \to Y$ is the blowing up of an $(\mathcal{I}, a)$-admissible $Z \subset Y$, with exceptional divisor E having ideal $\mathcal{I}_E$, then $\mathcal{I}\mathcal{O}_{Y'} = (\mathcal{I}_E)^a\, \mathcal{I}'$, with $\mathrm{maxord}(\mathcal{I}') \leq a$.

Order reduction is the following statement:

Theorem 2.5.1 (Order reduction)**.** *Let Y be a regular variety, $E_0 \subset Y$ a simple normal crossings divisor, and $\mathcal{I}$ an ideal sheaf, with*

$$\mathrm{maxord}(\mathcal{I}) = a.$$

There is a sequence of $(\mathcal{I}, a)$-admissible blowings up $Y_n \to Y_{n-1} \to \cdots \to Y_0 = Y$, with regular centers $Z_i \subset Y_i$ having normal crossings with E_i such that $\mathcal{I}\mathcal{O}_{X_n} = \mathcal{I}_n \mathcal{I}'_n$with $\mathcal{I}_n$ an invertible ideal supported on E_n and such that

$$\mathrm{maxord}(\mathcal{I}'_n) < a.$$

Order reduction implies principalization simply by induction on the maximal order $\mathrm{maxord}(\mathcal{I}) = a$: once $\mathrm{maxord}(\mathcal{I}'_n) = 0$ we have $\mathcal{I}\mathcal{O}_{X_n} = \mathcal{I}_n$ so only the exceptional part remains, which is supported on a simple normal crossings divisor by induction.

Hironaka himself used the *Hilbert–Samuel function*, an invariant much more refined than the order. It is a surprising phenomenon that resolution becomes easier to explain when one uses just the order, thus less information, see Encinas and Villamayor [2003].

It remains to prove order reduction.

3 Hironaka's method: order reduction

3.1 Differential operators. Nothing so far was particularly sensitive to the fact that we were working over $\mathbb{C}$, or even a field of characteristic 0. That starts changing now.

Since Y is regular, it has a *tangent bundle* T_Y. Local sections ∂ of T_Y are first order differential operators $\partial : \mathcal{O}_Y \to \mathcal{O}_Y$. As usual we denote the sheaf of sections of the tangent bundle with the same symbol T_Y, hoping the confusion can be overcome.

3.1.1 The characteristic 0 case. In characteristic 0, the sheaf of rings generated over $\mathcal{O}_Y$ by the operators in T_Y is the *sheaf of differential operators* $\mathfrak{D}_Y$. As a sheaf of $\mathcal{O}_Y$ modules it looks locally like the symmetric algebra $\mathrm{Sym}^\bullet(T_Y) = \oplus_{n\geq 0}\mathrm{Sym}^n(T_Y)$, but its ring structure is very different, as $\mathfrak{D}_Y$ is non-commutative. Still for any integer a there is a subsheaf $\mathfrak{D}_Y^{\leq a} \subset \mathfrak{D}_Y$ of differential operators of order $\leq a$, those sections which can be written in terms of monomials of order at most a in sections of T_Y. As a special case, one always has a splitting $\mathfrak{D}_Y^{\leq 1} = \mathcal{O}_Y \oplus T_Y$, the projection $\mathfrak{D}_Y^{\leq 1} \to \mathcal{O}_Y$ given by applying $\nabla \mapsto \nabla(1)$.

3.1.2 The general case. Things are quite different in characteristic $p > 0$: one can use the same definition, but in some sense it is deficient, because these differential operators do not detect pth powers. There is a natural and sophisticated replacement, which coincides with $\mathfrak{D}_Y$ in characteristic 0, and defined as follows:

On $Y \times Y$ consider the diagonal $\Delta \subset Y \times Y$. It is a closed subvariety, and one can consider its ideal $\mathcal{I}_\Delta$. The *sheaf of principal parts of order a of Y* is defined as $\mathcal{P}\mathcal{P}_Y^a = \mathcal{O}_{Y\times Y}/\mathcal{I}_\Delta^{a+1}$ - it is a sheaf of $\mathcal{O}_Y$-modules via either projection; its fiber at $p \in Y$ describes functions on Y up to order a at p. Its dual sheaf is $\mathfrak{D}^{\leq a} := (\mathcal{P}\mathcal{P}_Y^a)^\vee$, which *in characteristic 0* admits the concrete description given earlier. The natural projection $\mathcal{P}\mathcal{P}_Y^a \to \mathcal{P}\mathcal{P}_Y^{a-1}$ gives rise to an inclusion $\mathfrak{D}_Y^{\leq a-1} \subset \mathfrak{D}_Y^{\leq a}$, and one defines in general $\mathfrak{D}_Y = \cup_a \mathfrak{D}_Y^{\leq a}$.

This is nice enough, but the fact that in positive characteristics sections of $\mathfrak{D}_Y$ are not written as polynomials in sections of T_Y is the source of much trouble.

3.2 Derivatives and order. Let $\mathcal{I}$ be an ideal sheaf on Y and $y \in Y$ a point. Write $\mathcal{D}_Y^{\leq a}\mathcal{I}$ for the ideal generated by elements $\nabla(f)$, where ∇ an operator in $\mathcal{D}_Y^{\leq a}$ and f a section of $\mathcal{I}$. We have the following characterization:

$$\operatorname{ord}_y(\mathcal{I}) = \min\{a : (\mathcal{D}^{\leq a}\mathcal{I})_y = \mathcal{O}_{Y,y}\}.$$

In other words, the order of $\mathcal{I}$ at y is the minimum order of a differential operator ∇ such that for some $f \in \mathcal{I}_y$ the element $\nabla(f)$ does not vanish at y.

We define $\mathcal{T}(\mathcal{I}, a) := \mathcal{D}^{\leq a-1}\mathcal{I}$. In these terms, the set $V(\mathcal{I}, a)$ can be promoted to a *scheme*, the zero locus of an ideal: $V(\mathcal{I}, a) := V(\mathcal{T}(\mathcal{I}, a))$.[3]

This is not too surprising in characteristic 0 since we all learned calculus, but it may seem strange in characteristic $p > 0$. For instance, the order of (x^p) is p, since there is always an operator ∇ of order p such that $\nabla(x^p) = 1$. In characteristic 0 we can write

$$\nabla = \frac{1}{p!}\left(\frac{\partial}{\partial x}\right)^p,$$

but in characteristic p we have no such expression!

3.3 Induction and maximal contact hypersurfaces. We return to working over $\mathbb{C}$, in particular in characteristic 0, so we can use the letter p for a point of Y.

Remember that we want to prove order reduction of an ideal $\mathcal{I}$ of maximal order a. Hironaka's next idea was to use induction on dimension by restricting attention to a hypersurface H, in such a way that a suitable order reduction on H results, *by blowing up the same centers,* in order reduction of $\mathcal{I}$ on Y.

I am not being historically correct here, since Hironaka used invariants much more refined than order. I depart from history further, and use Giraud's concept of *maximal contact hypersurfaces*, adapted to orders rather than other invariants.

Definition 3.3.1. *Let $\mathcal{I}$ be an ideal of maximal order a.* A maximal contact hypersurface for $(\mathcal{I}, a)$ at p *is a hypersurface H* regular at p*, such that, in some neighborhood Y^0 of p we have $H \supseteq V(\mathcal{I}, a) = V(\mathcal{T}(\mathcal{I}, a))$, namely H contains the* scheme *of points where $\mathcal{I}$ has order a.*

3.4 Derivatives and existence in characteristic 0. It is not too difficult to show that *in characteristic 0*, a maximal contact hypersurface for $(\mathcal{I}, a)$ at p exists. Since $\mathcal{I}$ has maximal order a, we have $\mathcal{D}^{\leq a}(\mathcal{I}) = (1)$. Consider the ideal $\mathcal{T}(\mathcal{I}, a) = \mathcal{D}^{\leq a-1}(\mathcal{I})$. Since we are in characteristic 0, it must contain an antiderivative of 1, so $\operatorname{ord}_p(\mathcal{T}(\mathcal{I}, a)) \leq$

---

[3] This is the *right* scheme structure, as it satisfies an appropriate universal property.

1. Any local section x of $\mathcal{T}(\mathcal{I}, a)$ with order ≤ 1 gives a maximal contact hypersurface $\{x = 0\}$ at p.

Here is an example: suppose $\mathcal{I} = (f)$ where

$$f(x, y) = y^a + g_1(x)y^{a-1} + \cdots + g_{a-1}(x)y + g_a(x).$$

Then $\mathrm{ord}_{(0,0)}(f) = a$ exactly when $\mathrm{ord}_0(g_i) \geq i$ for all i. In characteristic 0 we may replace y by $y + g_1(x)/a$, so we may assume $g_1(x) = 0$. In this case $\partial^{a-1} f/\partial y^{a-1} = a! \cdot y$, so $\{y = 0\}$ is a maximal contact hypersurface at $(0, 0)$.

The definition I gave is pointwise. It is easy to see that if H is a maximal contact hypersurface for $(\mathcal{I}, a)$ at p then the same holds at any nearby p', so the concept is local. Unfortunately it is also not hard to cook up examples where there is no *global* maximal contact hypersurface which works everywhere. We will tackle this problem in Section 3.8 below.

3.4.1 Positive characteristics. Alas, there are fairly simple examples in charactersitic $p > 0$ where maximal contact hypersurfaces do not exist, see Narasimhan [1983] and Hauser [2010]. The whole discussion from here on simply does not work in characteristic > 0.

3.5 What should we resolve on H? For induction to work we need to decide what exactly we want to do on the hypersurface H. The example $\mathcal{I} = (f)$ above is instructive: just restricting $\mathcal{I}$ to H does not work!

Assume

$$f(x, y) = y^a + g_2(x)y^{a-2} + \cdots + g_{a-1}(x)y + g_a(x),$$

with $\mathrm{ord}_0(g_i) \geq i$ for all i. The restriction of $\mathcal{I}$ to the hypersurface $H = \{y = 0\}$ is the ideal $(g_a(x))$. Now clearly this ideal does not retain enough information from the original ideal. This is manifest with the notion of admissibility introduced in Section 2.5, as $(g_a(x), a)$-admissible centers in H will not always give $(\mathcal{I}, a)$-admissible centers in Y. For instance it might happen that $g_a = 0$, so every center on H is $(g_a(x), a)$-admissible, but $g_{a-1} \neq 0$, so not every center on H is $(\mathcal{I}, a)$-admissible on Y!

The collection of elements $g_2(x), \ldots, g_a(x)$ surely hold all the necessary information. However each comes with its own requirements: in order to reduce the order of $\mathcal{I}$ below a, we need to reduce the order of at least one of $g_i(x)$ below i.

We now generalize this discussion to arbitrary ideals.

3.5.1 Coefficient ideals and the induction scheme. In order to generalize this, we need to identify an analogue of these "elements" $g_i(x)$, and derivatives come to the rescue again. Let $\mathcal{I}$ be an ideal of maximal order a on a variety Y and H a maximal contact hypersurface. Then for any $i < a$ the ideals $\mathfrak{D}_Y^{\le i}\mathcal{I}$ have maximal order precisely $a-i$. It follows that the restricted ideals $(\mathfrak{D}_Y^{\le i}\mathcal{I})_{|H}$ have maximal order $\ge a-i$. These restrictions are the analogues of $g_{a-i}(x)$.

The following is at the technical core of the proof. I am aware of several proofs of this proposition, but they all seem to go a bit beyond the level of discussion I wish to maintain here. For reference, see Kollár [2007, Section 3.9].

Proposition 3.5.2. *Any sequence of $(\mathcal{I}, a)$-admissible blowings up has centers lying in H and its successive strict transforms. The resulting sequence of blowings up on H is $((\mathfrak{D}_Y^{\le i}\mathcal{I})_{|H}, a-i)$-admissible for every $i < a$.*

Conversely, every sequence of blowings up on H which is $((\mathfrak{D}_Y^{\le i}\mathcal{I})_{|H}, a-i)$-admissible for every $i < a$ gives rise, by blowing up the same centers on Y, to a sequence of $(\mathcal{I}, a)$-admissible blowings up.

Such an admissible sequence on H may be called an *order reduction* for the collection $((\mathfrak{D}_Y^{\le i}\mathcal{I})_{|H}, a-i)$ if it forms an order reduction for at least one of these pairs. It is a formal consequence of the proposition that an order reduction for $(\mathcal{I}, a)$ is the same as an order reduction for the collection $((\mathfrak{D}_Y^{\le i}\mathcal{I})_{|H}, a-i), i = 0, \dots, a$.

This may appear as troublesome: we wanted to prove order reduction for one ideal, and the induction requires us to prove order reduction for a collection of ideals. But there is a simple trick that allows one to replace this collection of ideals by a single ideal, in such a way that the notions of order reduction coincide:

Definition 3.5.3. *For an ideal $\mathcal{I}$ of maximal order a define its* coefficient ideal *to be the ideal sum $C(\mathcal{I}, a) := \sum(\mathfrak{D}_Y^{\le i}\mathcal{I})^{a!/(a-i)}$.*

Proposition 3.5.4 (Włodarczyk [2005, §3.4])**.** *Order reduction for $(C(\mathcal{I}, a)_{|H}, a!)$ is the same as order reduction for the collection $((\mathfrak{D}_Y^{\le i}\mathcal{I})_{|H}, a-i), i = 0, \dots, a$.*

We obtain:

Corollary 3.5.5 (Kollár [2007, Corollary 3.85])**.** *A sequence of blowings up is an order reduction for $(\mathcal{I}, a)$ if and only if it is an order reduction for $(C(\mathcal{I}, a)_{|H}, a!)$.*

3.6 A trouble of exceptional loci. I have been deliberately ignoring a subtle point. Theorem 2.5.1 about order reduction takes the additional datum of a divisor $E_0 \subset Y$. This is important for principalization, since once we reduce the order of $\mathcal{I}$ from a to $a-1$ with exceptional divisor say E_0, we want any further centers of blowing up used in further order reduction of $\mathcal{I}$ to have normal crossings with E_0.

For instance, in the example of a cuspidal curve above, the ideal $\mathcal{I} = x^2(z^2 - x)$ is of the form $\mathcal{I}_E^2 \mathcal{I}'$. The unique maximal contact hypersurface for $(\mathcal{I}', 1)$ is precisely X', the vanishing locus of $\mathcal{I}'$, but since it is tangent to E it does not have normal crossings with E.

The standard way to treat this is via a trick: one separates the relevant part of the ideal $\mathcal{I}'$ from the monomial part $\mathcal{I}_E$ by applying a suitable principalization for an ideal of the form $\mathcal{I}_E^\alpha + \mathcal{I}'^\beta$ describing the intersection of their loci. This is somewhat subtle and a bit disappointing. One feels that monomial ideals should only serve for good, as they are the goal.

I'll totally ignore this issue here, referring to Kollár [ibid., Section 3.13]. I have an excuse: in the procedure described below in my work with Temkin and Włodarczyk, this is not an issue at all, as monomial ideals become our best friends.

3.7 The problem of gluing. I postponed two important issues. Resolution of singularities requires *good* embedded resolution, *good* principalization, *good* order reduction: the process must be compatible with patching of open sets and independent of the embedding. A related issue is the fact that maximal contact hypersurfaces are not global, so patching open sets where maximal contact hypersurfaces do not overlap is required!

The classical approach has several ideas involved and has several levels of complexity. First, one devises a more elaborate *resolution invariant*, which records behavior of a given ideal on a sequence of nested maximal contact hypersurfaces. Second, one devises a class of transformations, called *test transformations*, which include admissible blowings up, restrictions to open sets, but also other operations, such as projections from a product.

One shows that ideals with the same invariant admit the same sequences of test transformations. It is a more subtle fact that the opposite is true - the invariant can be read off the test transformation. Once the dust settles it becomes clear that the order reduction one produces is independent of choices and is local, hence it can be patched along open sets. Also, the issue of the choice of embedding for resolution of singularities becomes local, hence it reduces to a simple principle I call *the re-embedding principle* in Section 3.10 below.

Some subset of this approach, in particular the fact that invariants can be read from the class of test transformations, is known as *Hironaka's trick*.

I have to admit that I never quite understood this approach until I read Włodarczyk's paper Włodarczyk [2005], which uses a completely different approach. After that transformative event I was able to read Bierstone and Milman [2008], and suddenly the classical approach was illuminated. I therefore prefer to present Włodarczyk's approach here, as perhaps others will experience the same transformation and subsequent illumination.

3.8 Włodarczyk's functoriality principle. Already Hironaka was interested in functorial properties of resolution of singularities. For instance, if X is a singular variety with a group G acting, ideally one would want the resolution to be G equivariant. Also if $X^0 \subset X$ is open, the resolution of X^0 should ideally be the restriction of that of X. This is stated explicitly in Bierstone and Milman [1997, §13].

Włodarczyk's great idea in Włodarczyk [2005] was that

> *functoriality is a powerful tool in the very proof of resolution of singularities.*

Moreover,

> *functoriality leads one to discover an order reduction algorithm.*

Włodarcyk requires one to take this very seriously. Indeed, for his principle to succeed one needs to use hidden symmetries, which are revealed only after $\mathcal{I}$ is tuned appropriately.

3.8.1 Smooth pullbacks. Let us first define the terms. Let $Y_n \to \cdots \to Y_0 = Y$ be an order reduction of $(\mathcal{I}, a)$ compatible with simple normal crossings divisor E. Let $Y' \to Y$ be a *smooth* morphism, what geometers call a *submersion*, such as an open embedding or a product with a regular variety. One can write $\mathcal{I}' = \mathcal{I}\mathcal{O}_Y$, $E' = E \times_Y Y'$ and $Y_i' = Y_i \times_Y Y'$, and then automatically $Y_n' \to \cdots \to Y_0' = Y'$ is an order reduction for $(\mathcal{I}', a)$, compatible with E', the *smooth pullback* order reduction. Some of the resulting steps might become trivial, in which case we drop them from the order reduction sequence.

Definition 3.8.2. *A* functorial order reduction *is a rule assigning to an ideal $\mathcal{I}$ on a regular variety Y with simple normal crossings divisor E and integer a such that* $\operatorname{maxord}(\mathcal{I}) \leq a$, *an order reduction $Y_n \to \cdots \to Y$, in such a way that for any smooth morphism $Y' \to Y$, the corresponding order reduction $Y_n' \to \cdots \to Y'$ is the smooth pullback order reduction of $Y_n \to \cdots \to Y$.*

To prove order reduction it suffices to produce functorial *order reduction on open patches, because then they automatically glue together.*

What about maximal contact hypersurfaces? Let us say we have produced functorial order reduction in dimension $\dim(Y) - 1$ and we wish to prove it for Y. We can choose a local maximal contact hypersurface $H \subset Y$ and reduce the order of the coefficient ideal $C(\mathcal{I}, a)_{|H}$. By Corollary 3.5.5 this results in *local* order reduction for $(\mathcal{I}, a)$, but *a priori* this depends on the choice of H. We claim that in fact there is no such dependence, and that the resulting order reduction is functorial on Y. For this we use $(\mathcal{I}, a)$-special automorphisms.

3.8.3 Special automorphisms. Let $Y_m \to \cdots \to Y_0$ be an $(\mathcal{I}, a)$-admissible sequence with centers $Z_i \subset Y_i$. Recall that this gives in particular a sequence of ideals $\mathcal{I}_i \subset \mathcal{O}_{Y_i}$

such that $Z_i \subset V(\mathcal{I}_i, a)$ and $\mathcal{I}_i \mathcal{O}_{Y_{i+1}} = \mathcal{I}_{E_{i+1}}^a \mathcal{I}_{i+1}$. An automorphism ϕ of Y is *special* if it *fixes* every $(\mathcal{I}, a)$-admissible sequence. This means that ϕ fixes $V(\mathcal{I}, a)$, in particular it fixes Z_0, hence it lifts to an automorphism ϕ_1 of Y_1, *which fixes* $V(\mathcal{I}_1, a)$, and inductively we obtain automorphisms ϕ_i of Y_i fixing $V(\mathcal{I}_i, a)$.

This is a very strong assumption on an automorphism, but Włodarczyk proved the following powerful result:

Proposition 3.8.4 (Włodarczyk [ibid.]). *Let $H_1, H_2 \subset Y$ be two local maximal contact hypersurfaces at $p \in V(\mathcal{I}, a)$. Then, after replacing Y by an étale neighborhood of p, there is a special automorphism ϕ of Y fixing p and sending H_1 to H_2.*

In particular, the functorial order reductions for $C(\mathcal{I}, a)_{|H_i}$ induce the same order reduction for $\mathcal{I}$, which is automatically functorial!

I deliberately did not require the automorphism ϕ to send $\mathcal{I}_i$ to itself, which would make the statement easier to grasp. Indeed, the following example shows that in general *it is impossible for ϕ to send $\mathcal{I}_i$ to itself,* and suggests that Proposition 3.8.4 is quite surprising and should require an ingenious idea.

Consider the ideal (xy) in the affine plane Y, with maximal order 2 attained at $V(\mathcal{I}, 2) = (0, 0)$, the origin. We have $\mathcal{D}^{\leq 2-1}\mathcal{I} = \mathcal{D}^{\leq 1}\mathcal{I} = (x, y)$, and so the lines $H_1 = \{x = 0\}$ and $H_2 = \{x + y = 0\}$ are both maximal contact hypersurfaces. Clearly any automorphism of Y sending H_1 to H_2 must change $\mathcal{I}$, since $\mathcal{I}_{|H_1} = 0$ and $\mathcal{I}_{|H_2} \neq 0$.

In this particular case the coefficient ideal is (x^2, xy, y^2), and the automorphism $(x, y) \mapsto (x + y, y)$ does send H_1 to H_2 fixing this ideal, so whatever procedure we apply using H_1 - in this case necessarily blowing up the origin - coincides with the process we apply using H_2.

3.8.5 Homogenization. The general case is slightly more subtle than the example: in general there is no automorphism carrying H_1 to H_2 fixing the coefficient ideal either. Searching for a natural replacement which is fixed under a special automorphism, Włodarcyk discovered the *homogenization* $\mathcal{H}(\mathcal{I}, a)$ described below.[4]

Definition 3.8.6. *Recall the notation $\mathcal{T}(\mathcal{I}, a) = \mathcal{D}_Y^{\leq a-1}\mathcal{I}$. The* homogenization *of $(\mathcal{I}, a)$ is the ideal*

$$\mathcal{H}(\mathcal{I}, a) := \sum_{i=0}^{a} \mathcal{D}^i(\mathcal{I})\, \mathcal{T}(\mathcal{I}, a)^i.$$

[4]A Different variant is used in Kollár [2007]; the treatment in Bravo, Garcia-Escamilla, and Villamayor U. [2012] in terms of differential Rees algebras provides a natural structure subsuming homogenization and coefficient ideals.

Włodarczyk's idea is that $\mathcal{I}$ lacks symmetries because it is not sufficiently tuned. In contrast, the ideal $\mathcal{H}(\mathcal{I}, a)$ is tuned to reveal the hidden symmetry ϕ.[5]

The ideal $\mathcal{H}(\mathcal{I}, a)$ is designed to contain all terms of Taylor expansions of elements of $\mathcal{I}$ in terms of any variable h in $\mathcal{T}(\mathcal{I}, a)$. If $H_1 = \{x = 0\}$, $H_2 = \{x + h = 0\}$ and $x = x_1, x_2, \ldots, x_m$ are local parameters of Y and p, chosen so that $x_1+h, x_2, \ldots, x_m$ also form local parameters, then the transformation $\phi(x_1, x_2, \ldots, x_m) = (x_1 + h, x_2, \ldots, x_m)$ is a local automorphism of Y formally sending $f(x_1, x_2, \ldots, x_m)$ to

$$\sum \frac{\partial^i f}{\partial x_1^i} h^i.$$

Note that $\frac{\partial^i f}{\partial x_1^i} h^i \in \mathcal{D}^i(\mathcal{I})\,\mathcal{T}(\mathcal{I}, a)^i$. Thus on formal completions this sends an element of $\mathcal{H}(\mathcal{I}, a)$ to an element of $\mathcal{H}(\mathcal{I}, a)$, and a bit of reflection shows that ϕ is a special automorphism with respect to $\mathcal{H}(\mathcal{I}, a)$. A standard argument allows to pass from completion to étale neighborhoods, hence ϕ defines a special automorphism with respect to $\mathcal{H}(\mathcal{I}, a)$ on a suitable étale neighborhood.

A simple computation shows:

Proposition 3.8.7. *Order reduction for $(\mathcal{I}, a)$ is equivalent to order reduction for $\mathcal{H}(\mathcal{I}, a)$.*

Thus ϕ is a special automorphism with respect to $(\mathcal{I}, a)$ as well!

3.9 A sketch of the algorithm. Let us summarize how one *functorially* reduces the order of a nonzero ideal $\mathcal{I}$ of maximal order $a > 0$ on a regular variety Y.

If $\dim(Y) = 0$ there is nothing to prove, since $\mathcal{I}$ is trivial hence of order 0. We assume proven order reduction in dimension $< \dim(Y)$.

We cover $V(\mathcal{I}, a)$ with open patches U possessing maximal contact hypersurfaces H_U. The coefficient ideal $C(\mathcal{I}, a)_{|H_U}$ has order $\geq a!$.

If this order is infinite, it means that $\mathcal{I}_{|U} = \mathcal{I}^a_{H_U}$, we simply blow up H_U and automatically the order of $\mathcal{I}$ is reduced on U.

Otherwise we can inductively reduce the order of this ideal by a functorial sequence of transformations $H_k \to \cdots \to H$ until the order drops below $a!$. By Corollary 3.5.5 these provide a local order reduction for $(\mathcal{I}, a)$ which patches together to a functorial order reduction by Proposition 3.8.4.

[5] In Bierstone and Milman [2008], Bierstone and Milman replace $(\mathcal{I}, a)$ by its equivalence class with respect to test transformation. With the "blurred vision" of equivalence classes of ideals, a hidden symmetry is again revealed. This is related to Hironaka's approach using the concept of *infinitely near points*.

3.10 The re-embedding principle. I still need to explain why a suitable embedded resolution of singularities implies resolution in general. If X is covered by open subsets X_i embedded in regular varieties Y_i, we want to claim that the resolutions $X_i' \to X_i$ agree on intersections $X_i \cap X_j$. Said another way, no matter how $X_i \cap X_j$ is embedded, the resolutions agree[6]. Since our procedures are functorial for étale maps and Y_i are regular, we may as well assume $Y_i = \mathbb{A}^{n_i}$. Finally affine spaces differ by iterated projections, so we are reduced to the following statement, which seems to follow from our procedures "by accident", see Kollár [2007, Claim 3.71.2]:

Proposition 3.10.1 (The re-embedding principle)**.** *Suppose $\mathcal{I}$ is an ideal on a regular variety Y. Consider the embedding $Y \subset Y_1 := Y \times \mathbb{A}^1$ sending $y \mapsto (y, 0)$. Let $\mathcal{I}_1 = \mathcal{I}\mathcal{O}_{Y_1} + (z)$, where z is the coordinate on $\mathbb{A}^1$. Then the principalization described above of $\mathcal{I}_1$ on Y_1 is obtained by taking the principalization of $\mathcal{I}$ on Y and blowing up the same centers, embedded in Y_1 and is transforms.*

I would very much like to say that this follows from functoriality, but this is not so simple (see Kollár's treatment). Instead, we look under the hood of principalization. To principalize $\mathcal{I}_1$ we need to reduce the order of $\mathcal{I}_1$ below 1. The order of $\mathcal{I}_1$ is 1, since z has order 1, and then z defines a maximal contact hypersurface, which is, seemingly by accident, precisely Y with the coefficient ideal $\mathcal{I}$, so the statement follows for the first blowing up. The local product structure persists after blowing up, so the statement holds for the entire order reduction procedure.

4 Toric varieties and toroidal embeddings

To proceed further it is useful to introduce a nice class of variety with "fairly simple" singularities.

4.1 Toric varieties. A *toric variety* is a *normal* variety X with a dense embedding $T = (\mathbb{C}^*)^n \hookrightarrow X$ such that the action of T on itself by translations extends to X. Here *normal* means that the local rings are integrally closed, a condition which guarantees that X is regular in codimension 1.

Toric varieties are a simple playing ground for algebraic geometers, as many aspects of a toric variety can be translated to combinatorics. To a toric variety X one associates a *fan* Σ_X, a collection of rational polyhedral cones in the lattice $N_T := \mathrm{Hom}(\mathbb{C}^*, T)$ which intersect similarly to cells of a CW complex. One makes toric varieties into a category on which arrows are torus equivariant morphisms which are surjective on the tori. Similarly

[6] There is a subtle issue of synchronization by codimension I will ignore. See Bierstone, Milman, and Temkin [2011, §5.3], Temkin [2011, §2.5.10]

fans form a category: a map of fans $\Sigma_1 \to \Sigma_2$ is induced by a map of lattices $N_1 \to N_2$ with finite cokernel, such that a cone of Σ_1 maps into a cone in Σ_2. There is an equivalence of categories

$$\{\text{toric varieties}\} \leftrightarrow \{\text{fans}\}.$$

A toric variety X is regular if and only if its fan Σ_X is regular: every cone is simplicial, and generators of its edges span a saturated lattice in N_T. Toric birational maps correspond to subdivisions of fans, and so toric resolution of singularities can be done by finding a regular subdivision, a fairly simple task.

There are great sources to learn the theory. See Kempf, Knudsen, Mumford, and Saint-Donat [1973], Oda [1988], and Fulton [1993].

4.2 Toroidal embeddings. All toric varieties are rational, so they have a limited chance to help with resolution of singularities. A *toroidal embedding* is an open embedding $U \subset X$ which locally analytically in the euclidean topology looks like a toric variety: for a point $p \in X$ there is a patch $V_p \subset X$ and a corresponding open set $W_p \subset Y$, where $T \subset Y$ is a toric variety, and an analytic isomorphism $V_p \to W_p$ carrying $V_p \cap U$ onto $W_p \cap T$.

One can speak of *toroidal morphisms* $X_1 \to X_2$ between toroidal embeddings: these are those morphisms which locally on the source look like toric morphisms of toric varieties.

As toroidal embeddings look locally like toric varieties their singularities are toric. It comes as no surprise that toric resolution of singularities extends quite easily to toroidal embeddings. In fact, one associates to a toroidal embedding $U \subset X$ a combinatorial gadget - a rational polyhedral cone complex - in a functorial manner. This is not an equivalence of categories, but it is still true that subdivisions correspond to toroidal birational morphisms, and the resolution procedure for fans extends to polyhedral cone complexes.

This is developed in Kempf, Knudsen, Mumford, and Saint-Donat [1973].

5 Resolution in families

I briefly recall known results on resolution in families, all relying on de Jong's *alteration method.*

5.1 The alteration theorem. In de Jong [1996], Johan de Jong discovered a method to replace a variety X by a regular variety X' with a morphism $X' \to X$ which is not necessarily birational, but is *proper, surjective and generically finite*. Such maps he called *alterations*, which differ from birational *modifications* in that the extension of function fields $K(X) \subset K(X')$ may be nontrivial.

Theorem 5.1.1 (de Jong [ibid.]). *Let X be a variety over a field of arbitrary characteristic and Z a subvariety. There is an alteration $f : X' \to X$ such that X' is smooth and $f^{-1}Z$ a simple normal crossings divisor.*

5.1.2 Sketch of proof. Here is the basic idea: assume for simplicity X is projective; blowing up makes Z into a divisor. One can choose a rational projection $X \dashrightarrow \mathbb{P}^{n-1}$, which becomes a morphism after replacing X by a modification, so that the generic fiber X_η over the generic point $\eta \in \mathbb{P}^n$ is a smooth curve, of some genus g, and Z can be viewed as a collection of k marked points on X_η. This corresponds to a morphism $\{\eta\} \to \overline{\mathfrak{M}}_{g,k}(X,d)$, the Kontsevich space of stable maps, where d is the degree of X_η with respect to some projective embedding. Properness of this moduli space provides us an alteration $B \to \mathbb{P}^n$ over which this extends to a family of stable maps $Y_0 \to X$ parametrized by B. Induction on the dimension allows us to assume that B is smooth and the degeneracy locus of Y_0/B is a simple normal crossings divisor. An inspection of Y_0 shows that it has the structure of toroidal embedding, hence admits a combinatorial resolution of singularities $Y \to Y_0$. The composite morphism $Y \to X$ is the required alteration.

5.2 Toroidalization. In the introduction, we stated resolution of singularities as the problem of making points of X *simple*. If instead we have a family of varieties $X \to B$ parametrized by a variety B, what should a resolution of singularities of the *family* mean? That is, when are the singularities of the family *simple*?

It is not hard to see that making all the fibers regular is impossible. Since we agree that toroidal singularities are rather simple, one might consider a toroidal morphism to represent a family with simple singlarities. Here are two solutions based on this idea:

Theorem 5.2.1 (Altered toroidalization, de Jong [ibid.]). *Let $X \to B$ be a dominant morphism of varieties. There are alterations $B_1 \to B$ and $X_1 \to X \times_B B_1$, with regular toroidal embedding structures $U_B \subset B_1$ and $U_X \subset X_1$, and a* toroidal *morphism $X_1 \to B_1$ making the following diagram commutative.*

(1)
$$\begin{array}{ccc} X_1 & \longrightarrow & X \\ \downarrow & & \downarrow \\ B_1 & \longrightarrow & B \end{array}$$

See improvement on this in Theorem 5.3.1 below. This is proven in much the same way as one proves de Jong's alteration theorem, Theorem 5.1.1. One needs to simply replace the projection $X \dashrightarrow \mathbb{P}^n$ by a relative projection $X \dashrightarrow \mathbb{P}^{d-1} \times B$, where d is the relative dimension of X over B.

In characteristic 0 one can inprove the situation, using *modifications* instead of *alterations*:

Theorem 5.2.2 (Toroidalization, Abramovich and Karu [2000, Theorem 2.1]). *Let $X \to B$ be a dominant morphism of complex projective varieties. There is a modification $B' \to B$, a modification $X' \to X$, and regular toroidal embedding structures $U_B \subset B'$ and $U_X \subset X'$, such that the map $X' \dashrightarrow B'$ is a* toroidal *morphism.*

Theorem 5.2.2 is proven using the following addition to de Jong's method, introduced in Abramovich and de Jong [1997]: in essence, one brings oneself to a situation as in equation (1), where the Galois group $Gal(K(X_1)/K(X))$ of the function field extension acts on the whole diagram. In characteristic 0 it turns out that the singularities of the quotients X_1/G and B_1/G can be resolved by toroidal methods - this is a feature of *tame* group actions in general. This sketch is only true in essense: in practice the Galois structure is intertwined with the inductive structure of the proof of Theorem 5.2.1.

Theorems 5.2.1 and 5.2.2 have two major disadvantages: they are by no means functorial, and they necessarily change the general fiber of $X \to B$ even if it is already regular.

5.3 Semistable reduction. We have already pointed out that toroidal embeddings can be resolved. The same is true to some extent for families as well. This means that the singularities in Theorems 5.2.1 and 5.2.2 can still be improved.

Let $X \to B$ be a toroidal morphism between toroidal embeddings. Assume X and B are regular. We say that $X \to B$ is *semistable* if locally at any point $x \in X$ there are distinct monomial variables $y_1, \ldots, y_k$ on B and $x_1, \ldots, x_m$ on X such that X is described by the following equations:

$$\begin{aligned} x_1 \cdots x_{l_1} &= y_1, \\ x_{l_1+1} \cdots x_{l_2} &= y_2, \\ \vdots \quad & \quad \vdots \\ x_{l_{k-1}+1} \cdots x_{l_k} &= y_k. \end{aligned}$$

This means that locally X is a product of families of the form $x_1 \cdots x_l = y$. This is truly the best one can hope for. A somewhat weaker and more flexible version would replace y_i by monomials m_i without common factors.

De Jong actually proved:

Theorem 5.3.1 (Altered semistable reduction, de Jong [1996]). *Let $X \to B$ be a dominant morphism of varieties over a field. There are alterations $B_1 \to B$ and $X_1 \to X \times_B B_1$,* *and regular toroidal embedding structures* $U_B \subset B_1$ *and* $U_X \subset X_1$*, with semistable morphism* $X_1 \to B_1$.

In characteristic 0 a somewhat weaker result was proven in Abramovich and Karu [2000, Theorem 0.3] where the generic fiber is modified but not altered. The strongest form is given in Karu [2000] for families of surfaces and threefolds:

Theorem 5.3.2 (Semistable reduction, Karu [ibid.]). *Let* $X \to B$ *be a dominant morphism of complex projective varieties with* $\dim(X) - \dim(B) \leq 3$*. There is an alteration* $B' \to B$*, a modification* $X' \to X \times_B B'$*, and regular toroidal embedding structures* $U_B \subset B'$ *and* $U_X \subset X'$*, such that the map* $X' \dashrightarrow B'$ *is semistable.*

The case of arbitrary relative dimension is conjectured in Abramovich and Karu [2000, Conjecture 0.2], and reduced to a completely combinatorial problem in Abramovich and Karu [ibid., Conjecture 8.4]. This remains open.

6 Resolution in toroidal orbifolds

6.1 Towards functorial resolution of families. All the toroidalization and semistable reduction theorems above suffer from severe non-functoriality, and most importantly they change the generic fiber even if it is smooth. This is a major drawback in application. For example, one would like to take a smooth family of varieties over an open base and compactify it with as simple fibers as possible. One envisions compactifying by closure in some projective space and applying toroidalization or semistable reduction. The theorems above do not provide this, as the original family is necessarily changed.

The approach I present here is to start from scratch and use Hironaka's method instead of de Jong's. People have thought of this for a while, notably Cutkosky, see Cutkosky [2005], though his goals are different.

6.2 Temkin's functoriality principle. Consider the family $X \to B$ where X is a regular surface with coordinates x, y and B a curve with coordinate t, and where the map is given by $xy = t$. This is a semistable family.

Now take the base change $B_1 \to B$ given by $s^2 = t$. The pullback family $X_1 \to B_1$ is given by equation $xy = s^2$, which is semistable in the weaker sense, as s^2 is a monomial. If we are to allow semistable families to be compatible with base change this additional flexibility is a must. From the point of view of resolution of singularities in families, both families are good, even though X_1 is singular.

An important point in this example is that $B_1 \to B$ and $X_1 \to X$ are *toroidal* morphisms. If we are to prove a functorial procedure for resolving singularities in families, the procedure must not modify families which are already semistable, so both $X \to B$ and $X_1 \to B_1$ must stay intact.

In Hironaka's resolution, the best way to ensure that regular varieties stay intact is to require the resolution to be functorial for smooth morphisms. Indeed if X is regular then $X \to \operatorname{Spec} \mathbb{C}$ is a smooth morphism, so the resolution of X must be the pullback of the resolution of $\operatorname{Spec} \mathbb{C}$, which is necessarily trivial.

In the semistable reduction problem, the best way to ensure that a toroidal $X \to B$ stays intact is to require functoriality for *toroidal* morphisms. For instance, if already $B = \operatorname{Spec} \mathbb{C}$ is just a point and X is toroidal, then the morphism $X \to \operatorname{Spec} \mathbb{C}$ is toroidal, so the procedure we produce for $X \to \operatorname{Spec} \mathbb{C}$ must be the pullback of the procedure we produce for the identity $\operatorname{Spec} \mathbb{C} \to \operatorname{Spec} \mathbb{C}$, which is necessarily trivial.

Temkin's functoriality principle is thus:

> *Toroidal morphisms form the smallest reasonable class of morphisms under which semistable reduction should be functorial.*

And, in view of Włodarczyk's philosophy,

> *functoriality for toroidal morphisms should lead one to discover a semistable reduction procedure.*

As a necessary prerequisite, *we must produce a resolution procedure, built on Hironaka's procedure, which is functorial for toroidal morphisms.*

6.3 Enter toroidal orbifolds. Having accepted Temkin's functoriality principle and agreed that we must produce a resolution procedure, built on Hironaka's procedure, which is functorial for toroidal morphisms, there is another surprising conclusion coming our way.

Temkin's principle forces us to take a departure from previous algorithms:

1. We can no longer work with a smooth ambient variety Y - *we must allow Y to have toroidal singularities.*

2. We cannot use only blowings up of smooth centers as our basic operations. Instead we use a class of modifications, called *Kummer blowings up*, which are stable under toroidal base change. These involve taking roots of monomials, in particular:

3. We can no longer work only with varieties Y - *we must allow Y to be a Deligne–Mumford stack.*

In essence, we are using "weighted blowings up on steroids". The stacks we need are as follows:

Definition 6.3.1. *A* toroidal orbifold (Y, U) *is a Deligne–mumford stack Y with diagonalizable inertia with a toroidal embedding $U \subset Y$.*

Item (3) may be hard to accept but it is absolutely essential for functoriality under toroidal morphisms.

Consider the affine plane Y, with coordinates x, u, where we endow the plane with a toroidal structure by declaring $U = Y \setminus \{u = 0\}$, so x is a parameter and u is a *monomial*. Say we want to principalize the ideal $\mathcal{I} := (x^2, u^2)$. Functoriality under smooth morphisms suggests that we must blow up the ideal (x, u). This indeed works and principalizes $\mathcal{I}$ in one step.

Now consider the affine plane Y_0, with coordinates x, v, similar to the above, but say we want to principalize the ideal $\mathcal{I}_0 := (x^2, v)$. Note that we have a toroidal morphism $Y \to Y_0$ given by $u^2 = v$, and $\mathcal{I} = \mathcal{I}_0 \mathcal{O}_Y$. Temkin's functoriality tells us that there should be a center on Y_0 whose pullback is (x, u). This center must therefore be defined by $(x, \sqrt{v})$!

There is only one way to deal with it, and that is to work systematically in a setup where one is allowed, when necessary, to take roots of monomials. This is possible precisely when working with toroidal orbifolds, as indicated above.

6.4 Principalization in toroidal orbifolds. In the joint work Abramovich, Temkin, and Włodarczyk [2017a] with Temkin and Włodarczyk we prove

Theorem 6.4.1 (Toroidal principalization)**.** *Let (Y, U) be a toroidal orbifold and $\mathcal{I}$ an ideal sheaf. There is a sequence $Y_n \to Y_{n-1} \to \cdots \to Y_0 = Y$ of Kummer blowings up, all supported over the vanishing locus $V(\mathcal{I})$, such that $\mathcal{I}\mathcal{O}_{Y_n}$ is an invertible monomial ideal.* The process is functorial for toroidal base change morphisms $Y' \to Y$.

6.5 Logarithmic derivatives and logarithmic orders. Temkin's functoriality principle suggests a natural replacement for derivatives.

In the toroidal world, the natural replacement for derivatives is provided by *logarithmic derivatives*: if u is a *monomial* function on a toroidal Y, then we use the operator $u\frac{\partial}{\partial u}$, which sends u to itself, but not $\frac{\partial}{\partial u}$. One then defines the sheaf $\mathcal{D}^{\leq a}_{Y,U}$ of *logarithmic* differential operators of order $\leq a$.

Given an ideal $\mathcal{I}$ on a toroidal (Y, U), one defines its *logarithmic order* at $p \in Y$ to be $\operatorname{logord}_p(\mathcal{I}) = \min\{a : \mathcal{D}^{\leq a}_{Y,U}(\mathcal{I}) = (1)\}$. This can take the value ∞ when the *monomial part* $\mathfrak{M}(\mathcal{I}) := \mathcal{D}^{\infty}_{Y,U}\mathcal{I}$ is nontrivial.

Then a miracle happens: using Temkin's functoriality and logarithmic derivatives, the broad outlines of principalization described above, as laid out in detail in Włodarczyk [2005], work in this new context once one has a stable formalism of toroidal orbifolds.

The formalism is developed in Abramovich, Temkin, and Włodarczyk [2017b] and the proof of the theorem is written out in Abramovich, Temkin, and Włodarczyk [2017a].

Our next task is to return to work on families of varieties. We hope to report on that in the near future.

References

D. Abramovich and A. J. de Jong (1997). "Smoothness, semistability, and toroidal geometry". *J. Algebraic Geom.* 6.4, pp. 789–801. MR: 1487237 (cit. on p. 560).

D. Abramovich and K. Karu (2000). "Weak semistable reduction in characteristic 0". *Invent. Math.* 139.2, pp. 241–273. MR: 1738451 (cit. on pp. 560, 561).

Dan Abramovich, Michael Temkin, and Jarosław Włodarczyk (Sept. 2017a). "Principalization of ideals on toroidal orbifolds". arXiv: 1709.03185 (cit. on pp. 563, 564).

– (Sept. 2017b). "Toroidal orbifolds, destackification, and Kummer blowings up". arXiv: 1709.03206 (cit. on p. 564).

José M. Aroca, Heisuke Hironaka, and José L. Vicente (1977). *Desingularization theorems*. Vol. 30. Memorias de Matemática del Instituto "Jorge Juan" [Mathematical Memoirs of the Jorge Juan Institute]. Consejo Superior de Investigaciones Científicas, Madrid. MR: 480502 (cit. on p. 543).

Edward Bierstone and Pierre D. Milman (1997). "Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant". *Invent. Math.* 128.2, pp. 207–302. MR: 1440306 (cit. on pp. 543, 554).

– (2008). "Functoriality in resolution of singularities". *Publ. Res. Inst. Math. Sci.* 44.2, pp. 609–639. MR: 2426359 (cit. on pp. 543, 553, 556).

Edward Bierstone, Pierre D. Milman, and Michael Temkin (2011). "$\mathbb{Q}$-universal desingularization". *Asian J. Math.* 15.2, pp. 229–249. MR: 2838221 (cit. on p. 557).

A. Bravo, M. L. Garcia-Escamilla, and O. E. Villamayor U. (2012). "On Rees algebras and invariants for singularities over perfect fields". *Indiana Univ. Math. J.* 61.3, pp. 1201–1251. MR: 3071698 (cit. on p. 555).

Bruno Buchberger (1965). "Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal". PhD thesis. Universitat of Innsbruck (cit. on p. 543).

Steven Dale Cutkosky (2004). *Resolution of singularities*. Vol. 63. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, pp. viii+186. MR: 2058431 (cit. on p. 543).

– (2005). "Local monomialization of transcendental extensions". *Ann. Inst. Fourier (Grenoble)* 55.5, pp. 1517–1586. MR: 2172273 (cit. on p. 561).

S. Encinas and O. Villamayor (1998). "Good points and constructive resolution of singularities". *Acta Math.* 181.1, pp. 109–158. MR: 1654779 (cit. on p. 543).

Santiago Encinas and Orlando Villamayor (2003). "A new proof of desingularization over fields of characteristic zero". In: *Proceedings of the International Conference on Algebraic Geometry and Singularities (Spanish) (Sevilla, 2001)*. Vol. 19. 2, pp. 339–353. MR: 2023188 (cit. on pp. 543, 549).

William Fulton (1993). *Introduction to toric varieties*. Vol. 131. Annals of Mathematics Studies. The William H. Roever Lectures in Geometry. Princeton University Press, Princeton, NJ, pp. xii+157. MR: 1234037 (cit. on p. 558).

Jean Giraud (1974). "Sur la théorie du contact maximal". *Math. Z.* 137, pp. 285–310. MR: 0460712 (cit. on p. 543).

A. Grothendieck (1971). "Travaux de Heisouké Hironaka sur la résolution des singularités". In: *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 1*. Gauthier-Villars, Paris, pp. 7–9. MR: 0414283 (cit. on p. 543).

Robin Hartshorne (1977). *Algebraic geometry*. Graduate Texts in Mathematics, No. 52. Springer-Verlag, New York-Heidelberg, pp. xvi+496. MR: 0463157 (cit. on p. 544).

Herwig Hauser (2006). "Seven short stories on blowups and resolutions". In: *Proceedings of Gökova Geometry-Topology Conference 2005*. Gökova Geometry/Topology Conference (GGT), Gökova, pp. 1–48. MR: 2282008 (cit. on p. 543).

– (2010). "On the problem of resolution of singularities in positive characteristic (or: a proof we are still waiting for)". *Bull. Amer. Math. Soc. (N.S.)* 47.1, pp. 1–30. MR: 2566444 (cit. on p. 551).

Heisuke Hironaka (1964). "Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II". *Ann. of Math. (2)* **79** *(1964), 109–203; ibid. (2)* 79, pp. 205–326. MR: 0199184 (cit. on p. 542).

A. J. de Jong (1996). "Smoothness, semi-stability and alterations". *Inst. Hautes Études Sci. Publ. Math.* 83, pp. 51–93. MR: 1423020 (cit. on pp. 558–560).

K. Karu (2000). "Semistable reduction in characteristic zero for families of surfaces and threefolds". *Discrete Comput. Geom.* 23.1, pp. 111–120. MR: 1727125 (cit. on p. 561).

G. Kempf, Finn Faye Knudsen, D. Mumford, and B. Saint-Donat (1973). *Toroidal embeddings. I*. Lecture Notes in Mathematics, Vol. 339. Springer-Verlag, Berlin-New York, pp. viii+209. MR: 0335518 (cit. on p. 558).

János Kollár (2007). *Lectures on resolution of singularities*. Vol. 166. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. vi+208. MR: 2289519 (cit. on pp. 543, 547, 552, 553, 555, 557).

R. Narasimhan (1983). "Monomial equimultiple curves in positive characteristic". *Proc. Amer. Math. Soc.* 89.3, pp. 402–406. MR: 715853 (cit. on p. 551).

Tadao Oda (1988). *Convex bodies and algebraic geometry*. Vol. 15. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. An introduction to the theory of toric varieties, Translated from the Japanese. Springer-Verlag, Berlin, pp. viii+212. MR: 922894 (cit. on p. 558).

Igor R. Shafarevich (2013). *Basic algebraic geometry. 1*. Third edition. Varieties in projective space. Springer, Heidelberg, pp. xiv+310. MR: 3100243 (cit. on p. 544).

Michael Temkin (2011). “Absolute desingularization in characteristic zero”. In: *Motivic integration and its interactions with model theory and non-Archimedean geometry. Volume II*. Vol. 384. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 213–250. MR: 2905858 (cit. on p. 557).

Orlando Villamayor (1989). “Constructiveness of Hironaka’s resolution”. *Ann. Sci. École Norm. Sup. (4)* 22.1, pp. 1–32. MR: 985852 (cit. on p. 543).

Jarosław Włodarczyk (2005). “Simple Hironaka resolution in characteristic zero”. *J. Amer. Math. Soc.* 18.4, pp. 779–822. MR: 2163383 (cit. on pp. 543, 552–555, 563).

Received 2017-11-15.

Department of Mathematics Box 1917
Brown University
Providence, RI 02912
USA

Dan Abramovich
abrmovic@math.brown.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (565–582)

POSITIVITY AND ALGEBRAIC INTEGRABILITY OF HOLOMORPHIC FOLIATIONS

Carolina Araujo

Abstract

The theory of holomorphic foliations has its origins in the study of differential equations on the complex plane, and has turned into a powerful tool in algebraic geometry. One of the fundamental problems in the theory is to find conditions that guarantee that the leaves of a holomorphic foliation are algebraic. These correspond to algebraic solutions of differential equations. In this paper we discuss algebraic integrability criteria for holomorphic foliations in terms of positivity of its tangent sheaf, and survey the theory of Fano foliations, developed in a series of papers in collaboration with Stéphane Druel. We end by classifying all possible leaves of del Pezzo foliations.

1 Introduction

The theory of holomorphic foliations has its origins in the study of differential equations on the complex plane $\mathbb{C}^2$. A central problem in this theory consists in finding conditions that guarantee the existence of algebraic solutions (Darboux [1878], Painlevé [1894], Poincaré [1891]). Consider for instance the following algebraic differential equations:

$$\frac{dy}{dx} = \frac{y}{x}, \tag{1-1}$$

$$\frac{dy}{dx} = y. \tag{1-2}$$

While the solutions to equation (1-1) are algebraic, namely $y = cx$, the solutions to equation (1-2) are mostly transcendental, namely $y = ce^x$. In both cases, the algebraic differential equation defines a saturated subsheaf $\mathcal{F} \subset T_{\mathbb{C}^2}$. By saturated we mean that $T_{\mathbb{C}^2}/\mathcal{F}$ is torsion free. We call this subsheaf a *foliation* of the plane. Curves that are everywhere

MSC2010: primary 37F75; secondary 14E30, 14M22.

tangent to $\mathscr{F}$ correspond to solutions of the equation, and are called *leaves of the foliation*. We remark that classically the word foliation refers to the partition of the plane into leaves. If we extend to $\mathbb{P}^2$ the foliations $\mathscr{F} \subset T_{\mathbb{C}^2}$ defined by the equations above, we obtain the saturated subsheaves $\mathscr{O}_{\mathbb{P}^2}(1) \subset T_{\mathbb{P}^2}$ in (1-1) and $\mathscr{O}_{\mathbb{P}^2} \subset T_{\mathbb{P}^2}$ in (1-2). As we shall see, the ampleness of $\mathscr{O}_{\mathbb{P}^2}(1)$ forces the solutions of equation (1-1) to be algebraic. This is the simplest manifestation of a series of results relating properties of positivity and algebraicity of holomorphic foliations.

In general, a *foliation* on a normal variety X is a saturated nonzero coherent subsheaf $\mathscr{F} \subsetneq T_X$ that satisfies the Frobenius integrability condition: $\mathscr{F}$ is closed under the Lie bracket. The Frobenius condition guarantees that a dense open subset of X is covered by analytic submanifolds whose tangent bundles are restrictions of $\mathscr{F}$. When these submanifolds are connected and maximal, we say that they are *leaves* of $\mathscr{F}$. We refer to Section 2 for definitions and generalities about holomorphic foliations on complex varieties, including notions of singularity for foliations.

In the last decades, foliations have proved to be a powerful tool in algebraic geometry. For instance, they play an important role in the proof of Green–Griffiths conjecture for surfaces of general type with positive Segre class (F. A. Bogomolov [1977], McQuillan [1998]). In many applications, a key problem is to find conditions that guarantee that a foliation has algebraic leaves, in which case we say that it is *algebraically integrable*, and to describe the structure of these algebraic subvarieties. We briefly mention two important instances of this.

1.1 (Miyaoka's criterion of uniruledness)**.** In Miyaoka [1987], Miyaoka proved a remarkable criterion of uniruledness in terms of numerical properties of the tangent sheaf. Namely, if X is a non-uniruled normal projective variety, then its cotangent sheaf Ω^1_X is generically semi-positive. This last condition means that for a sufficiently general complete intersection curve C on X, the restriction $(\Omega^1_X)_{|C}$ is semi-positive. This criterion in one of the main ingredients in the proof of abundance for threefolds (see Shepherd-Barron [1992]). Algebraic integrability of foliations plays a key role in Miyaoka's proof, which involves reduction to positive characteristic. Namely, if Ω^1_X is not generically semi-positive, then, using Harder–Narasimhan filtrations, one can construct a special foliation on X, whose restriction to a sufficiently general complete intersection curve C is ample. This foliation is shown to be algebraically integrable and covered by rational curves.

Miyaoka's algebraicity criterion has been extensively generalized. We mention Bost's arithmetic geometric counterpart (Bost [2001]), Bogomolov and McQuillan's criterion, which gives rationally connectedness of general leaves (see F. Bogomolov and McQuillan [2016] and also Kebekus, Solá Conde, and Toma [2007]), and most recently the extension by Campana and Paŭn, which considers positivity of the tangent sheaf with respect to more general movable curve classes (Campana and Păun [2015]). These criteria will be further discussed in Section 2.

1.2 (The structure of singular varieties with numerically trivial canonical class)**.** The Beauville–Bogomolov decomposition theorem asserts that, after étale cover, any compact Kähler manifold with numerically trivial canonical class is a product of a torus, Calabi–Yau and irreducible symplectic manifolds (Beauville [1983]). This structure result has been recently generalized to the singular setting in Druel [2017] and Höring and Peternell [2017]. Algebraic integrability of foliations plays a key role in the proof of this structure theorem. Namely, Greb, Kebekus, and Peternell [2016b] gives a decomposition of the tangent sheaf of a singular complex projective variety X with trivial canonical class into a direct sum of foliations with strong stability properties. Druel provides in Druel [2017] an algebraic integrability criterion to show that this decomposition of the tangent sheaf is induced by a product structure on a quasi-étale cover of X.

A common idea behind the algebraic integrability results for foliations discussed above is that positivity properties of foliations tend to increase algebraicity properties of the leaves. In a series of papers in collaboration with Stéphane Druel (Araujo and Druel [2013], Araujo and Druel [2014], Araujo and Druel [2016], Araujo and Druel [2017] and Araujo and Druel [2018]), we have investigated foliations with positive anticanonical class, which we call *Fano foliations*. For Fano foliations, a rough measure of positivity is the *index*. The index $\iota_{\mathcal{F}}$ of a Fano foliation $\mathcal{F}$ on a complex projective manifold X is the largest integer dividing $-K_{\mathcal{F}}$ in $\mathrm{Pic}(X)$. One special property of Fano foliations is that their leaves are always covered by rational curves, even when these leaves are not algebraic. Our works on Fano foliations with high index indicated that the higher is the index, the closer it is to being algebraically integrable. First of all, we have the following general bound on the index, in analogy with Kobayachi–Ochiai's theorem on the index of Fano manifolds:

Theorem 1.3 (Araujo, Druel, and Kovács [2008, Theorem 1.1])**.** *Let $\mathcal{F} \subsetneq T_X$ be a Fano foliation of rank r on a complex projective manifold X. Then $\iota_{\mathcal{F}} \leqslant r$, and equality holds only if $X \cong \mathbb{P}^n$.*

Foliations on $\mathbb{P}^n$ attaining the bound of Theorem 1.3 are induced by linear projections $\mathbb{P}^n \dashrightarrow \mathbb{P}^{n-r}$ (Déserti and Cerveau [2005, Théorème 3.8]). These results have been generalized to the singular setting in Araujo and Druel [2014] and Höring [2014].

Next we consider Fano foliations $\mathcal{F} \subsetneq T_X$ of rank r on complex projective manifolds X with index $\iota_{\mathcal{F}} = r - 1$. In analogy with the theory of Fano manifolds, we call them *del Pezzo foliations*. In contrast with the case of maximal index $\iota_{\mathcal{F}} = r$, there are examples of del Pezzo foliations on $\mathbb{P}^n$ with non-algebraic leaves. In fact, del Pezzo foliations on $\mathbb{P}^n$ were classified in Loray, Pereira, and Touzet [2013]. They are the following.

- A foliation induced by a dominant rational map $\mathbb{P}^n \dashrightarrow \mathbb{P}(1^{n-r}, 2)$, defined by $n-r$ linear forms and one quadric form, where $\mathbb{P}(1^{n-r}, 2)$ denotes the weighted projective space with weights $\underbrace{1, \dots, 1}_{r \text{ times}}, 2$.

- The linear pullback of a foliation $\mathscr{C}$ on $\mathbb{P}^{n-r+1}$ induced by a global vector field.

In the second case, if the vector field is general, then the leaves of $\mathscr{C}$ are transcendental, as the solutions to equation (1-2) above illustrates. Hence, the leaves of the del Pezzo foliation are not algebraic either. The following algebraic integrability result asserts that these are the only transcendental del Pezzo foliations on complex projective manifolds.

Theorem 1.4 (Araujo and Druel [2013, Theorem 1.1]). *Let $\mathscr{F}$ be a del Pezzo foliation on a complex projective manifold $X \not\cong \mathbb{P}^n$. Then $\mathscr{F}$ is algebraically integrable, and its general leaves are rationally connected.*

In the classical setting, del Pezzo manifolds were classified by Fujita in the 1980's. Most of them are complete intersections on weighted projective spaces. One may also expect a classification of del Pezzo foliations. For example, the only del Pezzo foliations on smooth quadric hypersurfaces are those induced by the restriction of linear projections from the ambient projective space (Araujo and Druel [2016, Proposition 3.18]). Moreover, quadrics are the only hypersurfaces that admit del Pezzo foliations (Araujo and Druel [2013, Corollary 4.8.]). We also know examples of del Pezzo foliations on certain Grassmannians and projective space bundles over projective spaces (Araujo and Druel [ibid., Sections 4 and 9]). In Section 3 we discuss the classification of del Pezzo foliations, under restrictions on the rank or on the singularities of the foliation. A complete classification of del Pezzo foliations seems to be a difficult problem. A step in this direction is a classification of all possible leaves of del Pezzo foliations, which is given in Proposition 3.3.

We remark that codimension 1 Fano foliations of large index on complex projective spaces have been classically studied. The *degree* d of a foliation $\mathscr{F}$ of rank r on $\mathbb{P}^n$ is defined as the degree of the locus of tangency of $\mathscr{F}$ with a general linear subspace $\mathbb{P}^{n-r} \subset \mathbb{P}^n$. It satisfies $d = r - \iota_{\mathscr{F}}$. So Fano foliations of large index on $\mathbb{P}^n$ are precisely those with small degree. Codimension 1 foliations on $\mathbb{P}^n$ of degree 0 and 1 were classified in Jouanolou [1979]. Those of degree 2 were classified in Cerveau and Lins Neto [1996].

Theorem 1.4 is in fact a special case of a more general result that gives a lower bound for the *algebraic rank* of a Fano foliation in terms of the index. The algebraic rank $r^a_{\mathscr{F}}$ of a foliation $\mathscr{F}$ on an algebraic variety X is the maximum dimension of an algebraic subvariety through a general point of X that is everywhere tangent to $\mathscr{F}$. A foliation is said to be *purely transcendental* if its algebraic rank is 0.

Theorem 1.5 (Araujo and Druel [2018, Corollary 1.6.]). *Let $\mathcal{F}$ be a Fano foliation of index $\iota_{\mathcal{F}}$ on a complex projective manifold X. Then $r^a_{\mathcal{F}} \geqslant \iota_{\mathcal{F}}$, and equality holds if and only if $X \cong \mathbb{P}^n$ and $\mathcal{F}$ is the pullback under a linear projection of a purely transcendental foliation on $\mathbb{P}^{n-r^a_{\mathcal{F}}}$ with trivial canonical class.*

Fano foliations may also play a distinguished role in the emerging theory of birational geometry of foliations. Higher dimensional algebraic geometry has had a strong influence in the study of holomorphic foliations. Techniques from the minimal model program have been successfully applied to the study of global properties of holomorphic foliations, leading to the birational classification of foliations by curves on surfaces (Brunella [1999], Mendes [2000], Brunella [2004]). In higher dimensions, very little is known and difficulties abound: Kawamata–Viehweg vanishing, abundance and resolution of singularities all fail in general for foliations. Positive results, mostly in dimension 3, include a reduction of singularities for foliations on threefolds (Cano [2004], McQuillan and Panazzolo [2013]) and a cone theorem for rank 2 foliations on threefolds (Spicer [2017]). We also mention that there are structure results for foliations with numerically trivial canonical class (Touzet [2008], Loray, Pereira, and Touzet [2013], Loray, Pereira, and Touzet [2011], Pereira and Touzet [2013]).

In the context of birational geometry of foliations, it is important to develop the theory of foliations on mildly singular varieties. In Section 2 we survey some aspects of the theory of foliations in this more general setup. In Section 3 discuss the classification of del Pezzo foliations on projective manifolds.

Notation and conventions. We always work over the field $\mathbb{C}$ of complex numbers. Varieties are always assumed to be irreducible. We denote by X_{ns} the nonsingular locus of a variety X. Given a sheaf $\mathcal{F}$ of $\mathcal{O}_X$-modules of generic rank r on a variety X, we denote by $\det(\mathcal{F})$ the sheaf $(\wedge^r \mathcal{F})^{**}$. If $\mathcal{G}$ is another sheaf of $\mathcal{O}_X$-modules on X, then we denote by $\mathcal{F}[\otimes]\mathcal{G}$ the sheaf $(\mathcal{F} \otimes \mathcal{G})^{**}$. When X is a normal variety, we denote by T_X the tangent sheaf $(\Omega^1_X)^*$.

Acknowledgments. I am deeply grateful to Stéphane Druel for the collaboration throughout the years and for his comments on earlier versions of this paper. I also thank Jorge Vitório Pereira for inspiring conversations about holomorphic foliations.

2 Foliations

In this section we define foliations on algebraic varieties, their canonical class and notions of singularities. We then discuss criteria of algebraic integrability and special properties of Fano foliations.

Definition 2.1. A *foliation* on a normal variety X is a saturated nonzero coherent subsheaf $\mathcal{F} \subsetneq T_X$ that is closed under the Lie bracket.

The *rank* r of $\mathcal{F}$ is the generic rank of $\mathcal{F}$.

The *canonical class* $K_{\mathcal{F}}$ of $\mathcal{F}$ is any Weil divisor on X such that $\mathcal{O}_X(-K_{\mathcal{F}}) \simeq \det(\mathcal{F})$.

We say that $\mathcal{F}$ is *Gorenstein* if $K_{\mathcal{F}}$ is Cartier.

We say that $\mathcal{F}$ is $\mathbb{Q}$-*Gorenstein* if $K_{\mathcal{F}}$ is $\mathbb{Q}$-Cartier.

Definition 2.2. Let $\mathcal{F}$ be a $\mathbb{Q}$-Gorenstein foliation of rank r on a normal variety X, and consider the induced map

$$\eta : \Omega_X^r = \wedge^r(\Omega_X^1) \to \wedge^r(T_X^*) \to \wedge^r(\mathcal{F}^*) \to \det(\mathcal{F}^*) \simeq \mathcal{O}_X(K_{\mathcal{F}}).$$

This is called a *Pfaff field* of rank r on X (Araujo and Druel [2014, Definition 3.4]). The *singular locus* S of $\mathcal{F}$ is the closed subscheme of X whose ideal sheaf $\mathcal{I}_S$ is the image of the associated map $\Omega_X^r[\otimes]\mathcal{O}_X(-K_{\mathcal{F}}) \to \mathcal{O}_X$. On the nonsingular locus X_{ns} of X, S_{red} consists of the points at which $\mathcal{F}_{|X_{\mathrm{ns}}}$ is not a subbundle of $T_{X_{\mathrm{ns}}}$. When $S = \varnothing$, we say that $\mathcal{F}$ is a *regular* foliation.

An analytic subvariety $Y \subset X$ is *invariant* under $\mathcal{F}$ if it is not contained in the singular locus of $\mathcal{F}$, and the restriction $\eta_{|Y_{\mathrm{ns}}} : \Omega^r_{X\,|Y_{\mathrm{ns}}} \to \mathcal{O}_X(K_{\mathcal{F}})_{|Y_{\mathrm{ns}}}$ factors through the natural map $\Omega^r_{X\,|Y_{\mathrm{ns}}} \to \Omega^r_{Y\,|Y_{\mathrm{ns}}}$.

A maximal invariant subvariety of dimension r is called a *leaf* of $\mathcal{F}$.

There are several notions of singularities for foliations. The notion of *reduced* foliations has been used in the birational classification of foliations by curves on surfaces (see Brunella [2004]). More recently, notions of singularities coming from the minimal model program have shown very useful when studying birational geometry of foliations. We introduce the notions of *canonical* and *log canonical* foliations following McQuillan [2008, Definition I.1.2]. Terminal and log terminal singularities can be defined analogously.

Definition 2.3. Let $\mathcal{F}$ be a $\mathbb{Q}$-Gorenstein foliation on a normal variety X. Let $\varphi : \tilde{X} \to X$ be a projective birational morphism. There is a unique foliation $\tilde{\mathcal{F}}$ on $\tilde{X}$ that agrees with $\varphi^*\mathcal{F}$ on the open subset of $\tilde{X}$ where φ is an isomorphism, and uniquely defined rational numbers $a(E, X, \mathcal{F})$'s such that

$$K_{\tilde{\mathcal{F}}} = \varphi^* K_{\mathcal{F}} + \sum_E a(E, X, \mathcal{F})E,$$

where E runs through all exceptional prime divisors for φ. As usual, the discrepancies $a(E, X, \mathcal{F})$'s do not depend on the birational morphism φ, but only on the valuations associated to the E's. We say that $\mathcal{F}$ is *canonical* if $a(E, X, \mathcal{F}) \geqslant 0$ for all E exceptional over X. We say that $\mathcal{F}$ is *log canonical* if $a(E, X, \mathcal{F}) \geqslant -\epsilon(E)$ for all E exceptional over X, where

$$\epsilon(E) = \begin{cases} 0 & \text{if } E \text{ is invariant by the foliation,} \\ 1 & \text{if } E \text{ is not invariant by the foliation.} \end{cases}$$

If a Gorenstein foliation is regular, then it is canonical (Araujo and Druel [2013, Lemma 3.10]).

Definition 2.4. Let X be a normal projective variety, and $\mathscr{F}$ a $\mathbb{Q}$-Gorenstein foliation on X. We say that $\mathscr{F}$ is a $\mathbb{Q}$-*Fano foliation* if $-K_{\mathscr{F}}$ is ample. In this case, the *index* of $\mathscr{F}$ is the largest positive rational number $\iota_{\mathscr{F}}$ such that $-K_{\mathscr{F}} \sim_{\mathbb{Q}} \iota_{\mathscr{F}} H$ for a Cartier divisor H on X.

If $\mathscr{F}$ is a $\mathbb{Q}$-Fano foliation of rank r on a normal projective variety X, then, by Höring [2014, Corollary 1.2], $\iota_{\mathscr{F}} \leqslant r$. Moreover, equality holds if and only if X is a generalized normal cone over a normal projective variety Z, and $\mathscr{F}$ is induced by the natural rational map $X \dashrightarrow Z$ (see also Araujo, Druel, and Kovács [2008, Theorem 1.1], and Araujo and Druel [2014, Theorem 4.11]). In Section 3 below we discuss Fano foliations of rank r and index $\iota_{\mathscr{F}} = r - 1$. We call these *del Pezzo* foliations.

Definition 2.5. Let $\mathscr{F}$ be a foliation on a normal variety X. We say that $\mathscr{F}$ is *algebraically integrable* if it is induced by a dominant rational map $\varphi\colon X \dashrightarrow Y$ with connected fibers into a normal variety. This means that, over the smooth locus $X^\circ \subset X$ of φ, we have $\mathscr{F}_{|X^\circ} = T_{X^\circ/Y}$.

In the setting of Definition 2.5, the general leaf of $\mathscr{F}$ is a general fiber of $\varphi_{|X^\circ}\colon X^\circ \dashrightarrow Y$. The map $\varphi\colon X \dashrightarrow Y$ is unique up to birational equivalence. It is often useful to take the variety Y to be the normalization of the unique proper subvariety of the Chow variety of X whose general point parametrizes the closure of a general leaf of $\mathscr{F}$ (viewed as a reduced and irreducible cycle in X). It comes with a universal cycle and induced morphisms:

(2-1)
$$\begin{array}{ccc} Z & \xrightarrow{\ \nu\ } & X \\ {\scriptstyle\pi}\downarrow & \swarrow{\scriptstyle \varphi}\ (\text{dashed}) & \\ Y. & & \end{array}$$

Here Z is normal, $\nu\colon Z \to X$ is birational and, for a general point $y \in Y$, $\nu\big(\pi^{-1}(y)\big) \subset X$ is the closure of a leaf of $\mathscr{F}$. We refer to the diagram (2-1) as the *family of leaves* of $\mathscr{F}$.

In our investigations of $\mathbb{Q}$-Gorenstein algebraically integrable foliations, it proved to be very useful to work with their *log leaves*, rather than with their leaves.

Definition 2.6 (Araujo and Druel [2014, Definition 3.11 and Remark 3.12])**.** Let $\mathscr{F}$ be a $\mathbb{Q}$-Gorenstein algebraically integrable foliation on a normal projective variety X. Let $i: F \to X$ be the normalization of the closure of a general leaf of $\mathscr{F}$. There is a canonically defined effective $\mathbb{Q}$-divisor Δ on F such that $K_F + \Delta \sim_{\mathbb{Q}} i^* K_{\mathscr{F}}$. If $\mathscr{F}$ is Gorenstein, then Δ is integral. The pair (F, Δ) is called a *general log leaf* of $\mathscr{F}$. Consider the family of leaves of $\mathscr{F}$ as in diagram (2-1). If $y \in Y$ is a general point, then $F \cong Z_y = \pi^{-1}(y)$. Over the smooth locus of X, we have $\operatorname{Supp}(\Delta) = \operatorname{Exc}(\nu) \cap Z_y$ under this identification, where $\operatorname{Exc}(\nu)$ denotes the exceptional locus of $\nu: Z \to X$ (Araujo and Druel [2018, Lemma 2.12]). In particular, over the smooth locus of X, $F \setminus \Delta$ is smooth.

For a Cartier divisor L on X, we write $L_{|F}$ for the pullback i^*L of L.

Proposition 2.7 (Araujo and Druel [ibid., Corollary 2.13])**.** *Let X be a smooth projective variety, and $\mathscr{F} \subsetneq T_X$ an algebraically integrable foliation on X, with general log leaf (F, Δ). Suppose that either $\rho(X) = 1$, or $\mathscr{F}$ is a Fano foliation. Then $\Delta \neq 0$.*

The following notion of log canonicity for algebraically integrable foliations is weaker than the notion introduced in Definition 2.3 (see Araujo and Druel [2013, Proposition 3.11]).

Definition 2.8. Let X be a normal projective variety, $\mathscr{F}$ a $\mathbb{Q}$-Gorenstein algebraically integrable foliation on X, and (F, Δ) its general log leaf. We say that $\mathscr{F}$ has *log canonical singularities along a general leaf* if (F, Δ) is log canonical.

The following is a special geometric property of algebraically integrable $\mathbb{Q}$-Fano foliation with log canonical singularities along a general leaf. It implies in particular that there is a common point in the closure of every general leaf.

Proposition 2.9 (Araujo and Druel [2016, Proposition 3.13])**.** *Let $\mathscr{F}$ be an algebraically integrable $\mathbb{Q}$-Fano foliation on a normal projective variety X, having log canonical singularities along a general leaf. Then there is a log canonical center of the general log leaf (F, Δ) whose image in X does not vary with the log leaf.*

Remark 2.10. The log canonicity assumption in Proposition 2.9 is necessary to guarantee the existence of a common point in the closure of a general leaf. For instance, consider the Grassmannian $\mathbb{G}(1, m)$ of lines on $\mathbb{P}^m$ for $m \geqslant 3$, and the rational map $\mathbb{G}(1, m) \dashrightarrow \mathbb{G}(1, m-1)$ induced by the projection $\mathbb{P}^m \dashrightarrow \mathbb{P}^{m-1}$ from a fixed point $P \in \mathbb{P}^m$. It induces a del Pezzo foliation $\mathscr{F}$ of rank 2 on $\mathbb{G}(1, m)$ whose general log leaf (F, Δ) is isomorphic to $(\mathbb{P}^2, 2\ell)$, where ℓ is a line in $\mathbb{P}^2$ (see Araujo and Druel [2013, Example 4.3]). More precisely, F is the $\mathbb{P}^2$ of lines contained in a plane $\Pi \cong \mathbb{P}^2$ of $\mathbb{P}^m$ that contains P, and ℓ is the line consisting of lines on Π through P. This log leaf is not log canonical, and there is no common point in the closure of a general leaf. Also, Araujo and Druel [ibid., Construction 9.10] produces del Pezzo foliations on projective space bundles over positive dimensional smooth projective varieties, which are contained in the relative tangent bundle of the fibration. Clearly there is no common point in the closure of a general leaf. The general log leaf in this case is isomorphic to a cone over $(C, 2P)$, where C is a rational normal curve and $P \in C$ is a point. Again, it is not log canonical.

It is useful to have effective algebraic integrability criteria for foliations. We recall Bogomolov and McQuillan's criterion (see also Bost [2001] and Kebekus, Solá Conde, and Toma [2007]).

Theorem 2.11 (F. Bogomolov and McQuillan [2016, Theorem 0.1])**.** *Let X be a normal projective variety, and $\mathcal{F}$ a foliation on X. Let $C \subset X$ be a complete curve disjoint from the singular loci of X and $\mathcal{F}$. Suppose that the restriction $\mathcal{F}_{|C}$ is an ample vector bundle on C. Then the leaf of $\mathcal{F}$ through any point of C is an algebraic variety, and the leaf of $\mathcal{F}$ through a general point of C is moreover rationally connected.*

This criterion can be applied to describe special properties of $\mathbb{Q}$-Fano foliations. Let X be a normal projective variety and $\mathcal{A}$ any ample line bundle on X. Consider the usual notions of slope and semi-stability with respect to $\mathcal{A}$ for torsion-free sheaves on X. Given a $\mathbb{Q}$-Fano foliation $\mathcal{F}$ of rank r on X, we have $\mu_{\mathcal{A}}(\mathcal{F}) = \frac{-K_{\mathcal{F}} \cdot \mathcal{A}^{n-1}}{r} > 0$. Let

$$0 = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_k = \mathcal{F} \tag{2-2}$$

be the Harder–Narasimhan filtration of $\mathcal{F}$ with respect to $\mathcal{A}$, with quotients $\mathcal{Q}_i = \mathcal{E}_i/\mathcal{E}_{i-1}$ satisfying $\mu_{\mathcal{A}}(\mathcal{Q}_1) > \mu_{\mathcal{A}}(\mathcal{Q}_2) > \ldots > \mu_{\mathcal{A}}(\mathcal{Q}_k)$. By the Mehta–Ramanathan Theorem, the Harder–Narasimhan filtration of $\mathcal{F}$ with respect to $\mathcal{A}$ commutes with restriction to a general complete intersection curve C. This generality conditions means that $C = H_1 \cap \cdots \cap H_{\dim(X)-1}$, where the H_i's are general members of linear systems $|m_i \mathcal{A}|$ for $m_i \in \mathbb{N}$ sufficiently large. It implies that each $\mathcal{F}_i$ is locally free along C. Set $s = \max\{i \geqslant 1 \mid \mu_{\mathcal{A}}(\mathcal{F}_i/\mathcal{F}_{i-1}) > 0\} \geqslant 1$. From the properties of the Harder–Narasimhan filtrations, it follows that $\mathcal{F}_i \subset T_X$ is a foliation for $1 \leqslant i \leqslant s$. From the slope conditions and properties of vector bundles on smooth curves (Hartshorne [1971, Theorem 2.4]), it follows that each restriction $(\mathcal{F}_i)_{|C}$ is ample. By Theorem 2.11, for $1 \leqslant i \leqslant s$, $\mathcal{F}_i \subset T_X$ is an algebraically integrable foliation, and the closure of a general leaf is rationally connected. This gives the following property of $\mathbb{Q}$-Fano foliations.

Corollary 2.12. *Let $\mathcal{F}$ be a $\mathbb{Q}$-Fano foliation on a normal projective variety X. Then $\mathcal{F}$ contains an algebraically integrable subfoliation whose general leaves are rationally connected.*

Remark 2.13. Let X be a Fano manifold with $\rho(X) = 1$ and consider the Harder–Narasimhan filtration of the tangent bundle T_X as in (2-2). Since $\rho(X) = 1$, any ample line bundle $\mathcal{A}$ gives the same notion of stability. A conjecture due to Iskovskikh predicts that T_X is (semi-)stable. If T_X is not semi-stable, the first nonzero subsheaf $\mathcal{F}_1$ in its Harder–Narasimhan filtration is called the maximal destabilizing subsheaf of T_X. Arguing as before, we see that $\mathcal{F}_1$ is an algebraically integrable Fano foliation on X. The slope inequality $\mu_{\mathcal{A}}(\mathcal{F}_1) > \mu_{\mathcal{A}}(T_X)$ is equivalent to the index inequality $\frac{\iota_{\mathcal{F}_1}}{\mathrm{rank}(\mathcal{F}_1)} > \frac{\iota_X}{\dim(X)}$. So in order to prove Iskovskikh's conjecture, one must rule out the existence of Fano foliations with large index on X.

More generally, one defines the *algebraic rank* $r^a_{\mathcal{F}}$ of a foliation $\mathcal{F}$ as the maximum dimension of an algebraic subvariety through a general point of X that is everywhere tangent to $\mathcal{F}$. If $\mathcal{F}$ has rank r, then $0 \leqslant r^a_{\mathcal{F}} \leqslant r$, and $r^a_{\mathcal{F}} = r$ if and only if $\mathcal{F}$ is algebraically integrable. When $r^a_{\mathcal{F}} = 0$, we say that the foliation $\mathcal{F}$ is *purely transcendental*. Suppose that $\mathcal{F}$ is not algebraically integrable. Then there exist a normal variety Y, unique up to birational equivalence, a dominant rational map with connected fibers $\varphi\colon X \dashrightarrow Y$, and a purely transcendental foliation $\mathcal{G}$ on Y such that $\mathcal{F}$ is the pullback of $\mathcal{G}$ via φ. This means that $\mathcal{F}_{|X^\circ} = (d\varphi^\circ)^{-1}(\mathcal{G}_{|Y^\circ})$, where $X^\circ \subset X$ and $Y^\circ \subset Y$ are smooth open subsets over which φ restricts to a smooth morphism $\varphi^\circ\colon X^\circ \to Y^\circ$.

We end this section by mentioning an algebraic integrability criterion of Campana and Păun, which generalizes Theorem 2.11. The classical notion of slope-stability with respect to an ample line bundle has been extended to allow stability conditions given by movable curve classes on $\mathbb{Q}$-factorial normal projective varieties (Campana and Peternell [2011], Greb, Kebekus, and Peternell [2016a], Campana and Păun [2015]). In this more general setting, one still has Harder–Narasimhan filtrations as in (2-2), although the analogous of the Mehta–Ramanathan Theorem fails in general. Let $\mathcal{F}$ be a foliation on a $\mathbb{Q}$-factorial normal projective variety X, and suppose that it has positive slope with respect to movable curve class $\alpha \in \mathrm{N}_1(X)_{\mathbb{R}}$. Consider the Harder–Narasimhan filtration of $\mathcal{F}$ with respect to α as in (2-2), and set $s = \max\{i \geqslant 1 \mid \mu_\alpha(\mathcal{F}_i/\mathcal{F}_{i-1}) > 0\} \geqslant 1$. Then the algebraic integrability criterion of Campana and Păun (Campana and Păun [ibid., Theorem 4.2]) implies that, for $1 \leqslant i \leqslant s$, $\mathcal{F}_i \subset T_X$ is an algebraically integrable foliation, and the closure of a general leaf is rationally connected. In particular, if $\mathcal{F}$ is a purely transcendental foliation, then $K_{\mathcal{F}}$ is pseudo-effective.

3 Classification of del Pezzo foliations

In this section we discuss classification results for del Pezzo foliations on projective manifolds.

Definition 3.1. A *del Pezzo foliation* is a Fano foliation $\mathcal{F}$ of rank $r \geqslant 2$ and index $\iota_{\mathcal{F}} = r - 1$.

In the Introduction we described all del Pezzo foliations on projective spaces and quadric hypersurfaces. We have the following classification of codimension 1 del Pezzo foliations on projective manifolds. See Araujo and Druel [2014, Theorem 1.3] for a more general statement.

Theorem 3.2. *Let $\mathcal{F} \subset T_X$ be a codimension 1 del Pezzo foliation on a smooth projective variety X. Then one of the following holds.*

1. *$X \cong \mathbb{P}^n$.*

2. *X is isomorphic to a quadric hypersurface in $\mathbb{P}^{n+1}$.*

3. *There is an inclusion of vector bundles $\mathcal{K} \subset \mathcal{E}$ on $\mathbb{P}^1$, inducing a relative linear projection*

$$\begin{array}{ccc} \mathbb{P}(\mathcal{E}) & \overset{\varphi}{\dashrightarrow} & \mathbb{P}(\mathcal{K}) \,, \\ & \searrow \quad \swarrow_{q} & \\ & \mathbb{P}^1 & \end{array}$$

such that $X \cong \mathbb{P}(\mathcal{E})$ and $\mathcal{F}$ is the pullback via φ of a foliation

$$q^*\Big(\det\big(\mathcal{E}/\mathcal{K}\big)\Big) \hookrightarrow T_{\mathbb{P}(\mathcal{K})} \,.$$

Moreover, one of the following holds.

- $(\mathcal{E}, \mathcal{K}) \cong \big(\mathcal{O}_{\mathbb{P}^1}(2) \oplus \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus 2}, \mathcal{O}_{p^1}(a)^{\oplus 2}\big)$ *for some positive integer a.*
- $(\mathcal{E}, \mathcal{K}) \cong \big(\mathcal{O}_{\mathbb{P}^1}(1)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus 2}, \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus 2}\big)$ *for some positive integer a.*
- $(\mathcal{E}, \mathcal{K}) \cong \big(\mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(a) \oplus \mathcal{O}_{\mathbb{P}^1}(b), \mathcal{O}_{\mathbb{P}^1}(a) \oplus \mathcal{O}_{\mathbb{P}^1}(b)\big)$ *for distinct positive integers a and b.*

Theorem 3.2 is the first instance of classification of del Pezzo foliations, when the ambient space is smooth and the codimension is 1. The classification problem can move in different directions. One may be interested in del Pezzo foliations on mildly singular varieties. In this direction, Araujo and Druel [ibid., Theorem 1.3] allows X to be factorial and canonical. The conclusion is the same as in Theorem 3.2, with the additional possibility of X being a cone over certain surfaces of Picard rank 1. One may be interested in classifying codimension 1 Fano foliations of slightly smaller index. Fano foliations $\mathcal{F}$ of rank $r \geq 3$ and index $\iota_{\mathcal{F}} = r - 2$ are called *Mukai foliations*. In Araujo and Druel [2017], we have classified codimension 1 Mukai foliations on projective manifolds. Finally, one is often interested in del Pezzo foliations of arbitrary rank. For the rest of this paper, we consider del Pezzo foliations of arbitrary rank on projective manifolds. Recall from Theorem 1.4

that, except when the ambient space is $\mathbb{P}^n$, del Pezzo foliations are always algebraically integrable. As a step in the classification problem, we give a classification of all possible general log leaves of del Pezzo foliations.

Proposition 3.3. *Let $\mathscr{F}$ be an algebraically integrable del Pezzo foliation of rank $r \geqslant 2$ on a smooth projective variety X, with general log leaf (F, Δ). Let L be an ample divisor on X such that $-K_{\mathscr{F}} \sim (r-1)L$. Then $\big(F, \Delta, L_{|F}\big)$ satisfies one of the following conditions.*

1. $\big(F, \mathscr{O}_F(\Delta), \mathscr{O}_F(L_{|F})\big) \cong (\mathbb{P}^r, \mathscr{O}_{\mathbb{P}^r}(2), \mathscr{O}_{\mathbb{P}^r}(1))$.
2. *(F, Δ) is a cone over (Q^m, H), where Q^m is a smooth quadric hypersurface in $\mathbb{P}^{m+1}$ for some $2 \leqslant m \leqslant r$, $H \in \big|\mathscr{O}_{Q^m}(1)\big|$, and $L_{|F}$ is a hyperplane under this embedding.*
3. $\big(F, \mathscr{O}_F(\Delta), \mathscr{O}_F(L_{|F})\big) \cong (\mathbb{P}^2, \mathscr{O}_{\mathbb{P}^2}(1), \mathscr{O}_{\mathbb{P}^2}(2))$.
4. *$\big(F, \mathscr{O}_F(L_{|F})\big) \cong (\mathbb{P}_{\mathbb{P}^1}(\mathscr{E}), \mathscr{O}_{\mathbb{P}(\mathscr{E})}(1))$, and one of the following holds:*
   - *(a) $\mathscr{E} = \mathscr{O}_{\mathbb{P}^1}(1) \oplus \mathscr{O}_{\mathbb{P}^1}(d)$ for some $d \geqslant 2$, and $\Delta \sim_{\mathbb{Z}} \sigma + f$, where σ is the minimal section and f a fiber of $\mathbb{P}(\mathscr{E}) \to \mathbb{P}^1$.*
   - *(b) $\mathscr{E} = \mathscr{O}_{\mathbb{P}^1}(2) \oplus \mathscr{O}_{\mathbb{P}^1}(d)$ for some $d \geqslant 2$, and Δ is a minimal section.*
   - *(c) $\mathscr{E} = \mathscr{O}_{\mathbb{P}^1}(1) \oplus \mathscr{O}_{\mathbb{P}^1}(1) \oplus \mathscr{O}_{\mathbb{P}^1}(d)$ for some $d \geqslant 1$, and $\Delta = \mathbb{P}_{\mathbb{P}^1}(\mathscr{O}_{\mathbb{P}^1}(1) \oplus \mathscr{O}_{\mathbb{P}^1}(1))$.*
5. *(F, Δ) is a cone over (C_d, B), where C_d is rational normal curve of degree d in $\mathbb{P}^d$ for some $d \geqslant 2$, $B \in \big|\mathscr{O}_{\mathbb{P}^1}(2)\big|$, and $L_{|F}$ is a hyperplane under this embedding.*
6. *(F, Δ) is a cone over the pair (4a) above, and $L_{|F}$ is a hyperplane section of the cone.*

Proof. By Proposition 2.7, $\Delta \neq 0$, and so $K_F + (r-1)L_{|F} \sim -\Delta$ is not pseudo-effective.

Let $\nu\colon \tilde{F} \to F$ be a resolution of singularities, and set $\tilde{L} = \nu^* L_{|F}$. Then $\tilde{L}$ is nef and big. In the language of Andreatta [2013], $(\tilde{F}, \tilde{L})$ is a quasi-polarized variety. Moreover, $K_{\tilde{F}} + (r-1)\tilde{L}$ is not pseudo-effective. As in the proof of Höring [2014, Lemma 2.5], we run a $(K_{\tilde{F}} + (r-1)\tilde{L})$-MMP, $\varphi : \tilde{F} \dashrightarrow F'$. Since $K_{\tilde{F}} + (r-1)\tilde{L}$ is not pseudo-effective, it ends with a Mori fiber space $F' \to Z$:

$$\begin{array}{ccccc} (\tilde{F}, \tilde{L}) & \dashrightarrow & (F_i, L_i) & \dashrightarrow & (F', L'). \\ \big\downarrow{\scriptstyle \nu} & & & & \\ F & & & & \end{array}$$

By Andreatta [2013, Proposition 3.6], if (F_i, L_i) is an r-dimensional terminal $\mathbb{Q}$-factorial quasi-polarized variety, and $\mathbb{R}_{\geqslant 0}[C]$ is a $(K_{F_i} + (r-1)L_i)$-negative extremal ray of birational type, then $L_i \cdot C = 0$. Therefore, $\varphi : \tilde{F} \dashrightarrow F'$ is an MMP relative to F, and there exists a morphism $\nu' \colon F' \to F$ such that $\nu = \nu' \circ \varphi$. In particular, $L' = (\nu')^* L_{|F}$ is nef and big. Quasi-polarized varieties (F', L') with a Mori fiber space structure induced by a $(K_{F'} + (r-1)L')$-negative extremal ray were classified in Andreatta [ibid., Proposition 3.5]. They are the following:

(a) $(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(1))$.

(b) $(Q^r, \mathcal{O}_{Q^r}(1))$, where Q^r is a quadric hypersurface in $\mathbb{P}^{r+1}$.

(c) A cone over $(\mathbb{P}^2, \mathcal{O}_{\mathbb{P}^2}(2))$ (where the vertex is allowed to be empty).

(d) $(\mathbb{P}_B(\mathscr{E}), \mathcal{O}_{\mathbb{P}_B(\mathscr{E})}(1))$, where $\mathscr{E}$ is a nef and big vector bundle of rank r over a smooth curve B.

In case (a) we have $F' \cong F$ and $\Delta \in \left|\mathcal{O}_{\mathbb{P}^r}(2)\right|$.

In case (b) we have $F' \cong F$ and $\Delta \in \left|\mathcal{O}_{Q^r}(1)\right|$. Moreover, since $F \setminus \Delta$ is smooth, (F, Δ) is a cone over (Q^m, H), where Q^m is a smooth quadric hypersurface in $\mathbb{P}^{m+1}$ and $H \in \left|\mathcal{O}_{Q^m}(1)\right|$, for some $1 \leqslant m \leqslant r$. When $m = 1$, F is isomorphic to a cone over a conic curve, and this case will be covered under case (d) below.

In case (c), we have $F' \cong F$, (F, Δ) is a cone over the Veronese embedding of $(\mathbb{P}^2, \ell)$ in $\mathbb{P}^5$. Here ℓ is a line in $\mathbb{P}^2$ and thus Δ is a cone over a smooth conic. In particular, (F, Δ) is log canonical and Δ is its only log canonical center. By Proposition 2.9, the image of Δ does not vary with the log leaf. Suppose that the vertex V of (F, Δ) is nonempty. Then the image of V does not vary with the log leaf either. Therefore any point of X can be connected to any point in the image of V in X by a rational curve of L-degree 1. This implies that $X \cong \mathbb{P}^n$. From the classification of del Pezzo foliations on $\mathbb{P}^n$, we see that this is not possible. We conclude that $V = \varnothing$ and $\big(F, \mathcal{O}_F(\Delta), \mathcal{O}_F(L_{|F})\big) \cong (\mathbb{P}^2, \mathcal{O}_{\mathbb{P}^2}(1), \mathcal{O}_{\mathbb{P}^2}(2))$.

We now consider case (d). Denote by g is the genus of B and by f a fiber of the natural morphism $\pi : F' \to B$. Write ξ for a divisor on F' such that $\mathcal{O}_{\mathbb{P}(\mathscr{E})}(1) \cong \mathcal{O}_{F'}(\xi)$, and e for a divisor on B such that $\mathcal{O}_B(e) \cong \det \mathscr{E}$. Then $-K_{F'} = r\xi + \pi^*(-e - K_B)$.

We consider the following possibilities:

(d-1) $F' \cong F$.

(d-2) $\nu' \colon F' \to F$ is the divisorial contraction induced by ξ.

(d-3) $\nu'\colon F' \to F$ is the small contraction induced by ξ.

In case (d-1), since F is rationally connected, we must have $B \cong \mathbb{P}^1$ and $\mathcal{E}$ is an ample vector bundle on $\mathbb{P}^1$. Write

$$\mathcal{E} = \mathcal{O}_{\mathbb{P}^1}(a_1) \oplus \cdots \oplus \mathcal{O}_{\mathbb{P}^1}(a_r),$$

with $1 \leqslant a_1 \leqslant \cdots \leqslant a_r$. We have $0 \neq \Delta \in \left|\xi + (-\sum a_i + 2) f\right|$, and so

$$1 \leqslant h^0\big(\mathbb{P}(\mathcal{E}), \mathcal{O}_{\mathbb{P}(\mathcal{E})}(1) \otimes \pi^*\mathcal{O}_{\mathbb{P}^1}(-\textstyle\sum a_i + 2)\big) = h^0\big(\mathbb{P}^1, \mathcal{E}(-\textstyle\sum a_i + 2)\big).$$

This implies that $(r; a_1, \ldots, a_{r-1}) \in \big\{(2;1), (2;2), (3;1,1)\big\}$. When $\mathcal{E} = \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(1)$, F is a smooth quadric surface, and has already been consider. So we have one of the possibilities described in (4).

In case (d-2), F is $\mathbb{Q}$-factorial and $(\nu')^*\Delta \in \left|\mathcal{O}_{\mathbb{P}(\mathcal{E})}(1) \otimes \pi^*(\det \mathcal{E}^\vee \otimes \omega_B^\vee)\right|$, and so

$$1 \leqslant h^0\big(\mathbb{P}(\mathcal{E}), \mathcal{O}_{\mathbb{P}(\mathcal{E})}(1) \otimes \pi^*(\det \mathcal{E}^\vee \otimes \omega_B^\vee)\big) = h^0\big(B, \mathcal{E} \otimes \det \mathcal{E}^\vee \otimes \omega_B^\vee\big).$$

By Hartshorne's theorem (Hartshorne [1971, Theorem 2.4]), $\mathcal{E}$ is nef if and only if it has no quotient with negative slope. So we must have $g \in \{0, 1\}$. Moreover, if $g = 1$, then $\det \mathcal{E}$ is the first nonzero piece of the Harder–Narasimhan filtration of $\mathcal{E}$, and $\mathcal{Q} = \mathcal{E}/\det \mathcal{E}$ is a vector bundle on B. The only member of $\left|\mathcal{O}_{\mathbb{P}(\mathcal{E})}(1) \otimes \pi^*(\det \mathcal{E}^\vee)\right|$ is precisely the projectivization of $\mathcal{Q}$. It is the exceptional divisor of ν', which is impossible. So we conclude that $B \cong \mathbb{P}^1$, F is a cone over a rational normal curve of degree d for some $d \geqslant 1$, and $L_{|F}$ is a hyperplane under this embedding. When $d = 1$, we have $F \cong \mathbb{P}^r$. So we may assume that $d \geqslant 2$. A straightforward computation shows that Δ is linearly equivalent to two times a ruling of the cone.

In case (d-3), $\mathcal{O}_F(L_{|F})$ pulls back to $\mathcal{O}_{\mathbb{P}(\mathcal{E})}(1)$, and Δ is the image under the small contraction of a nonzero effective divisor $\Delta' \in \left|\mathcal{O}_{\mathbb{P}(\mathcal{E})}(1) \otimes \pi^*(\det \mathcal{E}^\vee \otimes \omega_B^\vee)\right|$. As before, $g \in \{0, 1\}$. Moreover, if $g = 1$, then $\det \mathcal{E}$ is the first nonzero piece of the Harder–Narasimhan filtration of $\mathcal{E}$, $\mathcal{Q} = \mathcal{E}/\det \mathcal{E}$ is a vector bundle on B, and $\Delta' \cong \mathbb{P}(\mathcal{Q})$ is the only member of $\left|\mathcal{O}_{\mathbb{P}(\mathcal{E})}(1) \otimes \pi^*(\det \mathcal{E}^\vee)\right|$. In particular, (F, Δ) is log canonical and Δ is its only log canonical center. By Proposition 2.9, the image of Δ does not vary with the log leaf, and so the image of its singular locus V does not vary with the log leaf either. Note that V is the image of the exceptional locus of ν'. Therefore any point of X can be connected to any point of V by a rational curve of L-degree 1, and thus $X \cong \mathbb{P}^n$. From the classification of del Pezzo foliations on $\mathbb{P}^n$, we see that this is not possible. So we conclude that $B \cong \mathbb{P}^1$. As in (d-1), we see that one of the following holds:

- $\mathcal{E} = \mathcal{O}_{\mathbb{P}^1}^{\oplus r-2} \oplus \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(d)$ for some $d \geqslant 2$, $r > 2$, and $\Delta' \sim_{\mathbb{Z}} \sigma + f$, where $\sigma = \mathbb{P}_{\mathbb{P}^1}\big(\mathcal{O}_{\mathbb{P}^1}^{\oplus r-2} \oplus \mathcal{O}_{\mathbb{P}^1}(1)\big)$, and f a fiber of $\mathbb{P}_{\mathbb{P}^1}(\mathcal{E}) \to \mathbb{P}^1$.
- $\mathcal{E} = \mathcal{O}_{\mathbb{P}^1}^{\oplus r-2} \oplus \mathcal{O}_{\mathbb{P}^1}(2) \oplus \mathcal{O}_{\mathbb{P}^1}(d)$ for some $d \geqslant 2$, $r > 2$, and $\Delta = \mathbb{P}_{\mathbb{P}^1}\big(\mathcal{O}_{\mathbb{P}^1}^{\oplus r-2} \oplus \mathcal{O}_{\mathbb{P}^1}(2)\big)$.
- $\mathcal{E} = \mathcal{O}_{\mathbb{P}^1}^{\oplus r-3} \oplus \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(d)$ for some $d \geqslant 1$, $r > 3$, and $\Delta = \mathbb{P}_{\mathbb{P}^1}(\mathcal{O}_{\mathbb{P}^1}^{\oplus r-3} \oplus \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(1))$.

The first case can also be described as a cone over the pair (4a) above, yielding (6). The latter two cases can be described as cones over the pairs (4b) and (4c) above, respectively. In these cases, (F, Δ) is log canonical and Δ is its only log canonical center. Moreover, Δ is a cone with vertex $V \neq \varnothing$ over a conic and a smooth quadric surface, respectively. By Proposition 2.9, the image of Δ does not vary with the log leaf, and so the image of V does not vary with the log leaf either. Therefore any point of X can be connected to any point in the image of V in X by a rational curve of L-degree 1, and thus $X \cong \mathbb{P}^n$. From the classification of del Pezzo foliations on $\mathbb{P}^n$, we see that this is not possible. □

We do not know examples of del Pezzo foliations with log leaves of type (3) and (6) described in Proposition 3.3. When the general log leaf of $\mathcal{F}$ is log canonical, Proposition 2.9 may be used to recover the ambient space X. For example, consider case (1), when $\big(F, \mathcal{O}_F(\Delta), \mathcal{O}_F(L_{|F})\big) \cong (\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(2), \mathcal{O}_{\mathbb{P}^r}(1))$. If (F, Δ) is log canonical, then Proposition 2.9 yields a common point $x \in X$ in the closure of a general leaf. Therefore any point of X can be connected to x by a rational curve of L-degree 1. This implies that $X \cong \mathbb{P}^n$. On the other hand, there are del Pezzo foliations with general log leaf of type (1) and not log canonical (see Remark 2.10).

We end this paper by reviewing a classification of del Pezzo foliations on projective manifolds under restrictions on the singularities of the foliation $\mathcal{F}$. Namely, we assume that $\mathcal{F}$ has log canonical singularities and is locally free along a general leaf. Recall from Theorem 1.4 that, if $X \ncong \mathbb{P}^n$, then del Pezzo foliations on X are always algebraically integrable. If we remove the log canonicity assumption, we know more examples of del Pezzo foliations, as discussed in Remark 2.10. One may be able to remove the locally freeness assumption using the classification of log leaves in Proposition 3.3.

Theorem 3.4. *[Araujo and Druel [2013, 9.1 and Theorems 1.3, 9.2, 9.6] and Araujo and Druel [2016, Theorem 1.3]] Let $\mathcal{F}$ be an algebraically integrable del Pezzo foliation of rank r on a projective manifold X. Suppose that $\mathcal{F}$ has log canonical singularities and is locally free along a general leaf. Then one of the following holds.*

1. *$r = 2$ and $\rho(X) = 1$.*

2. *$X \cong \mathbb{P}^n$.*

3. *X is isomorphic to a quadric hypersurface in $\mathbb{P}^{n+1}$.*

4. *$X \cong \mathbb{P}^1 \times \mathbb{P}^k$, $r \in \{2,3\}$ and $\mathcal{F}$ is the pullback via the second projection of a foliation on $\mathbb{P}^k$ induced by a linear projection.*

5. *There is an inclusion of vector bundles $\mathcal{K} \subset \mathcal{E}$ on $\mathbb{P}^1$, inducing a relative linear projection*

$$\begin{array}{ccc} \mathbb{P}(\mathcal{E}) & \overset{\varphi}{\dashrightarrow} & \mathbb{P}(\mathcal{K}) \,, \\ & \searrow \quad \swarrow_{q} & \\ & \mathbb{P}^1 & \end{array}$$

such that $X \cong \mathbb{P}(\mathcal{E})$ and $\mathcal{F}$ is the pullback via φ of a foliation

$$q^*\Big(\det\big(\mathcal{E}/\mathcal{K}\big)\Big) \hookrightarrow T_{\mathbb{P}(\mathcal{K})} \,.$$

Moreover, one of the following holds.

- *$(\mathcal{E},\mathcal{K}) \cong \big(\mathcal{O}_{\mathbb{P}^1}(2) \oplus \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus m}, \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus m}\big)$ for some $a \geqslant 1$ and $m \geqslant 2$ ($r = 2$).*
- *$(\mathcal{E},\mathcal{K}) \cong \big(\mathcal{O}_{\mathbb{P}^1}(1)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus m}, \mathcal{O}_{\mathbb{P}^1}(a)^{\oplus m}\big)$ for some $a \geqslant 1$ and $m \geqslant 2$ ($r = 3$).*
- *$\mathcal{E} \cong \mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{K}$, where $\mathcal{K}$ is an ample vector bundle not isomorphic to $\mathcal{O}_{\mathbb{P}^1}(a)^{\oplus m}$ for any integer a ($r = 2$).*

6. *There is an inclusion of vector bundles $\mathcal{K} \subset \mathcal{E}$ on $\mathbb{P}^k$, with $k \geqslant 2$ and $\mathcal{E}/\mathcal{K} \cong \mathcal{O}_{\mathbb{P}^k}(1)$, inducing a relative linear projection*

$$\begin{array}{ccc} \mathbb{P}(\mathcal{E}) & \overset{\varphi}{\dashrightarrow} & \mathbb{P}(\mathcal{K}) \,, \\ & \searrow \quad \swarrow_{q} & \\ & \mathbb{P}^k & \end{array}$$

such that $X \cong \mathbb{P}(\mathcal{E})$ and $\mathcal{F}$ is the pullback via φ of a foliation $q^*\mathcal{O}_{\mathbb{P}^k}(1) \hookrightarrow T_{\mathbb{P}(\mathcal{K})}$ ($r = 2$).

References

M. Andreatta (2013). "Minimal model program with scaling and adjunction theory". *Internat. J. Math.* 24.2, pp. 1350007, 13. MR: 3045341 (cit. on pp. 576, 577).

C. Araujo and S. Druel (2013). "On Fano foliations". *Adv. Math.* 238, pp. 70–118 (cit. on pp. 567, 568, 571, 572, 579).

– (2014). "On codimension 1 del Pezzo foliations on varieties with mild singularities". *Math. Ann.* 360.3-4, pp. 769–798 (cit. on pp. 567, 570–572, 575).

– (2016). "On Fano foliations 2." English. In: *Foliation theory in algebraic geometry. Proceedings of the conference, New York, NY, USA, September 3–7, 2013*. Cham: Springer, pp. 1–20 (cit. on pp. 567, 568, 572, 579).

– (2017). "Codimension 1 Mukai foliations on complex projective manifolds". *J. Reine Angew. Math.* 727, pp. 191–246 (cit. on pp. 567, 575).

– (2018). "Characterization of generic projective space bundles and algebraicity of foliations". To appear in Commentarii Mathematici Helvetici. arXiv: 1711.10174 (cit. on pp. 567, 569, 572).

C. Araujo, S. Druel, and S. J. Kovács (2008). "Cohomological characterizations of projective spaces and hyperquadrics". *Invent. Math.* 174.2, pp. 233–253 (cit. on pp. 567, 571).

A. Beauville (1983). "Variétés Kähleriennes dont la première classe de Chern est nulle". *J. Differential Geom.* 18.4, 755–782 (1984). MR: 730926 (cit. on p. 567).

F. A. Bogomolov (1977). "Families of curves on a surface of general type". *Dokl. Akad. Nauk SSSR* 236.5, pp. 1041–1044. MR: 0457450 (cit. on p. 566).

F. Bogomolov and M. McQuillan (2016). "Rational curves on foliated varieties". In: *Foliation theory in algebraic geometry*. Simons Symp. Springer, Cham, pp. 21–51 (cit. on pp. 566, 573).

J.-B. Bost (2001). "Algebraic leaves of algebraic foliations over number fields". *Publ. Math. Inst. Hautes Études Sci.* 93, pp. 161–221 (cit. on pp. 566, 573).

M. Brunella (1999). "Minimal models of foliated algebraic surfaces". *Bull. Soc. Math. France* 127.2, pp. 289–305. MR: 1708643 (cit. on p. 569).

– (2004). *Birational geometry of foliations*. IMPA Mathematical Publications. Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, pp. iv+138 (cit. on pp. 569, 570).

F. Campana and M. Păun (2015). "Foliations with positive slopes and birational stability of orbifold cotangent bundles". arXiv: 1508:0245 (cit. on pp. 567, 574).

F. Campana and T. Peternell (2011). "Geometric stability of the cotangent bundle and the universal cover of a projective manifold". *Bull. Soc. Math. France* 139.1. With an appendix by Matei Toma, pp. 41–74 (cit. on p. 574).

F. Cano (2004). "Reduction of the singularities of codimension one singular foliations in dimension three". *Ann. of Math. (2)* 160.3, pp. 907–1011. MR: 2144971 (cit. on p. 569).

D. Cerveau and A. Lins Neto (1996). "Irreducible components of the space of holomorphic foliations of degree two in $\mathbb{C}\mathrm{P}(n)$, $n \geq 3$". *Ann. of Math. (2)* 143.3, pp. 577–612 (cit. on p. 568).

G. Darboux (1878). "Mémoire sur les équations différentielles algébriques du premier ordre et du premier degré". fre. *Bulletin des Sciences Mathématiques et Astronomiques* 2.1, pp. 151–200 (cit. on p. 565).

J. Déserti and D. Cerveau (2005). "Feuilletages et actions de groupes sur les espaces projectifs". *Mém. Soc. Math. Fr. (N.S.)* 103, vi+124 pp. (2006) (cit. on p. 567).

S. Druel (2017). "A decomposition theorem for singular spaces with trivial canonical class of dimension at most five". To appear in Invent. Math. (cit. on p. 567).

D. Greb, S. Kebekus, and T. Peternell (2016a). "Movable curves and semistable sheaves". *Internat. Math. Res. Notices* 2016.2, pp. 536–570 (cit. on p. 574).

– (2016b). "Singular spaces with trivial canonical class". In: *Minimal models and extremal rays (Kyoto, 2011)*. Vol. 70. Adv. Stud. Pure Math. Math. Soc. Japan, [Tokyo], pp. 67–113. MR: 3617779 (cit. on p. 567).

R. Hartshorne (1971). "Ample vector bundles on curves". *Nagoya Math. J.* 43, pp. 73–89 (cit. on pp. 573, 578).

A. Höring (2014). "Twisted cotangent sheaves and a Kobayashi-Ochiai theorem for foliations". *Ann. Inst. Fourier (Grenoble)* 64.6, pp. 2465–2480 (cit. on pp. 567, 571, 576).

A. Höring and T. Peternell (2017). "Algebraic integrability of foliations with numerically trivial canonical bundle". arXiv: 1710.06183 (cit. on p. 567).

J. P. Jouanolou (1979). *Équations de Pfaff algébriques*. Vol. 708. Lecture Notes in Mathematics. Berlin: Springer, pp. v+255 (cit. on p. 568).

S. Kebekus, L. Solá Conde, and M. Toma (2007). "Rationally connected foliations after Bogomolov and McQuillan". *J. Algebraic Geom.* 16.1, pp. 65–81 (cit. on pp. 566, 573).

F. Loray, J. V. Pereira, and F. Touzet (2011). "Singular foliations with trivial canonical class". arXiv: 1107.1538 (cit. on p. 569).

– (2013). "Foliations with trivial canonical bundle on Fano 3-folds". *Math. Nachr.* 286.8-9, pp. 921–940 (cit. on pp. 567, 569).

M. McQuillan (1998). "Diophantine approximations and foliations". *Inst. Hautes Études Sci. Publ. Math.* 87, pp. 121–174. MR: 1659270 (cit. on p. 566).

– (2008). "Canonical models of foliations". *Pure Appl. Math. Q.* 4.3, part 2, pp. 877–1012 (cit. on p. 570).

M. McQuillan and D. Panazzolo (2013). "Almost étale resolution of foliations". *J. Differential Geom.* 95.2, pp. 279–319. MR: 3128985 (cit. on p. 569).

L. G. Mendes (2000). "Kodaira dimension of holomorphic singular foliations". *Bol. Soc. Brasil. Mat. (N.S.)* 31.2, pp. 127–143. MR: 1785264 (cit. on p. 569).

Y. Miyaoka (1987). "Deformations of a morphism along a foliation and applications". In: *Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985)*. Vol. 46. Proc. Sympos. Pure Math. Providence, RI: Amer. Math. Soc., pp. 245–268 (cit. on p. 566).

P. Painlevé (1894). *Leçons, sur la théorie analytique des équations différentielles: professées à Stockholm (septembre, octobre, novembre 1895) sur l'invitation de S. M. le roi de Suède et de Norwège*. Leçons, sur la théorie analytique des équations différentielles: professées à Stockholm (septembre, octobre, novembre 1895) sur l'invitation de S. M. le roi de Suède et de Norwège v. 11, no. 4. A. Hermann (cit. on p. 565).

J. V. Pereira and F. Touzet (2013). "Foliations with vanishing Chern classes". *Bull. Braz. Math. Soc. (N.S.)* 44.4, pp. 731–754. MR: 3167130 (cit. on p. 569).

H. Poincaré (1891). "Sur l'intégration algébrique des équations différentielles du premier ordre et du premier degré". *Rendiconti del Circolo Matematico di Palermo* 5, pp. 161–191 (cit. on p. 565).

N. I. Shepherd-Barron (1992). "Miyaoka's Theorem on the Generic Seminegativity of T_X and on the Kodaira Dimension of Minimal Regular Threefolds". In: *Flips and Abundance for Algebraic Threefolds. A summer seminar at the University of Utah, Salt Lake City, 1991*. Vol. 211. J. Kollár, editor. Asterisque, pp. 103–114 (cit. on p. 566).

C. Spicer (2017). "Higher dimensional foliated Mori theory". arXiv: 1709.06850 (cit. on p. 569).

F. Touzet (2008). "Feuilletages holomorphes de codimension un dont la classe canonique est triviale". *Ann. Sci. Éc. Norm. Supér. (4)* 41.4, pp. 655–668 (cit. on p. 569).

Received 2018-01-04.

Carolina Araujo
IMPA
Estrada Dona Castorina 110
Rio de Janeiro 22460-320 Brazil
caraujo@impa.br

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (583–608)

BIRATIONAL GEOMETRY OF ALGEBRAIC VARIETIES

CAUCHER BIRKAR

Abstract

This is a report on some of the main developments in birational geometry in recent years focusing on the minimal model program, Fano varieties, singularities and related topics, in characteristic zero.

1 Introduction

It is not a comprehensive survey of all advances in birational geometry, e.g. we will not touch upon the positive characteristic case which is a very active area of research. We will work over an algebraically closed field k of characteristic zero. Varieties are all quasi-projective.

Birational geometry, with the so-called minimal model program at its core, aims to classify algebraic varieties up to birational isomorphism by identifying "nice" elements in each birational class and then classifying such elements, e.g study their moduli spaces. Two varieties are birational if they contain isomorphic open subsets. In dimension one, a nice element in a birational class is simply a smooth and projective element. In higher dimension though there are infinitely many such elements in each class, so picking a representative is a very challenging problem. Before going any further lets introduce the canonical divisor.

1.1 Canonical divisor. To understand a variety X one studies subvarieties and sheaves on it. Subvarieties of codimension one and their linear combinations, that is, divisors play a crucial role. Of particular importance is the canonical divisor K_X. When X is smooth this is the divisor (class) whose associated sheaf $\mathcal{O}_X(K_X)$ is the canonical sheaf $\omega_X := \det \Omega_X$ where Ω_X is the sheaf of regular differential forms. When X is only normal, K_X is the closure of the canonical divisor of the smooth locus. In general, the canonical divisor is the only special non-trivial divisor attached to X. It plays an important

MSC2010: primary 14E30; secondary 14J45, 14B05, 14C20, 14E05, 14E07.

role in algebraic geometry, e.g. in duality theory and Riemann-Roch formula, and also in differential and arithmetic geometry. It is a central object in birational geometry.

Example. Assume $X = \mathbb{P}^d$. Then $K_X \sim -(d+1)H$ where $H \subset \mathbb{P}^d$ is a hyperplane.

Example. Assume $X \subset \mathbb{P}^d$ is a smooth hypersurface of degree r. Then we have $K_X \sim (-d-1+r)H|_X$ where $H \subset \mathbb{P}^d$ is a hyperplane not containing X.

Example. If X is a toric variety, then $K_X \sim -\Lambda$ where Λ is the sum of the torus-invariant divisors.

1.2 Varieties with special canonical divisor. Let X be a projective variety with "good" singularities (by this we mean klt or lc singularities defined below, see Section 2.4).

$$\text{We say } X \text{ is } \begin{cases} \textit{Fano} & \text{if } K_X \text{ is anti-ample} \\ \textit{Calabi-Yau} & \text{if } K_X \text{ is numerically trivial} \\ \textit{canonically polarised} & \text{if } K_X \text{ is ample} \end{cases}$$

Note that here we consider Calabi-Yau varieties in a weak sense, that is, we do not require the vanishing $h^i(X, \mathcal{O}_X) = 0$ for $0 < i < \dim X$ which is usually assumed in other contexts. For example, abelian varieties are Calabi-Yau by our definition.

The special varieties just defined are of great importance in algebraic geometry (e.g. birational geometry, moduli theory, derived categories), differential geometry (e.g. Kähler-Einstien metrics, stability), arithmetic geometry (e.g. existence and density of rational points), and mathematical physics (e.g. string theory and mirror symmetry). They behave much better compared to a randomly chosen variety.

Example. Assume X is a smooth projective curve of genus g. If $g = 0$, then $X \simeq \mathbb{P}^1$ which is Fano. If $g = 1$, then X is an elliptic curve, hence a Calabi-Yau. If $g \geq 2$, then X is canonically polarised.

Example. Assume $X \subset \mathbb{P}^d$ is a smooth hypersurface of degree r. If $r \leq d$, then X is Fano. If $r = d + 1$, then X is Calabi-Yau. If $r > d + 1$, then X is canonically polarised.

1.3 Minimal model program. Now we give a brief description of the minimal model program (MMP). Pick a variety W. Using resolution of singularities we can modify W so that it is smooth and projective. However, being smooth and projective is not very special as in dimension at least two these properties are shared by infinitely many other varieties in the same birational class. It is then natural to look for a more special representative. One of the main aims of birational geometry is to show that we can dismantle W birationally and reconstruct it using canonically polarised, Calabi-Yau, and Fano varieties. To be more precise we want to establish the following conjecture formulated in its simplest form.

Conjecture 1.4 (Minimal model and abundance)**.** *Each variety W is birational to a projective variety Y with "good" singularities such that either*

- *Y is canonically polarised, or*
- *Y admits a Fano fibration, or*
- *Y admits a Calabi-Yau fibration.*

In particular, even if W is smooth, Y may be singular. In fact singularity theory is an indispensable part of modern birational geometry.

As the name suggests the conjecture actually consists of two parts, the *minimal model conjecture* and the *abundance conjecture*. The minimal model conjecture essentially says that we can find Y such that K_Y is *nef* meaning K_Y intersects every curve non-negatively, or else there is a K_Y-negative fibration $Y \to Z$ which means we have a *Fano fibration*. The abundance conjecture essentially says that if Y is not canonically polarised and if it does not admit a Fano fibration, then it admits a K_Y-trivial fibration $Y \to Z$ which means we have a *Calabi-Yau fibration*. The minimal model conjecture holds in dimension ≤ 4 by Mori [1988], Shokurov [1993], Kawamata [1992c], and Shokurov [2003, 2009] in full generality, and in any dimension for varieties of general type by Birkar, Cascini, Hacon, and McKernan [2010] while the abundance conjecture is proved in dimension ≤ 3 by Miyaoka [1988] and Kawamata [1992a], and in any dimension for varieties of general type by Shokurov [1985] and Kawamata [1984] (also see Birkar [2012] and references therein for more results). We should also mention that the *non-vanishing conjecture* which is a special case of (a suitable reformulation of) the abundance conjecture implies the minimal model conjecture by Birkar [2010, 2011].

Given a smooth projective W, how can we get to Y? This is achieved via *running the MMP* which is a step by step program making the canonical divisor K_W more positive by successively removing or replacing curves along which K_W is not positive. It gives a (conjecturally finite) sequence of birational transformations

$$W = W_1 \dashrightarrow W_2 \dashrightarrow \cdots \dashrightarrow W_n = Y$$

consisting of *divisorial contractions*, *flips*, and a last step canonically trivial contraction. The required contractions and flips exist by Shokurov [1985] and Kawamata [1984] and Birkar, Cascini, Hacon, and McKernan [2010] and Hacon and McKernan [2010]. An important ingredient is the *finite generation* of the k-algebra

$$R = \bigoplus_{m \geq 0} H^0(W, mK_W)$$

in its various forms; see Birkar, Cascini, Hacon, and McKernan [2010], Hacon and McKernan [2010], and Shokurov [2003].

A serious issue with the MMP is that we do not know whether it actually stops at some step W_n. What is not clear is if the MMP can produce an infinite sequence of flips. In other words, the minimal model conjecture is reduced to the following.

Conjecture 1.5 (Termination)**.** *There is no infinite sequence of flips.*

The two-dimensional case of the MMP is classical developed in the early 20th century by Castelnuovo, Enriques, etc. The three-dimensional case (in characteristic zero) was developed in the 70's-90's through work of many people notably Iitaka, Iskovskikh, Kawamata, Kollár, Mori, Reid, Shokurov, Ueno, etc. The higher dimensional case is still conjectural but a large portion of it has been established since the turn of the century by many people including Birkar, Cascini, Hacon, M^c^Kernan, Shokurov, Xu, etc, involving many difficult problems of local and global nature.

1.6 Pluricanonical systems, Kodaira dimension and Iitaka fibration. Let W be a smooth projective variety. The space of sections $H^0(W, mK_W)$, for $m \in \mathbb{Z}$, and their associated linear systems $|mK_W|$ are of great importance. When W is one-dimensional the linear system $|K_W|$ determines its geometry to a large extent. Indeed the genus g of W is just $h^0(W, K_W)$ which is encoded in $|K_W|$. Moreover, if $g \geq 2$, then $|K_W|$ is base point free, and if in addition W is not hyperelliptic, then $|K_W|$ defines an embedding of X into a projective space of dimension $g-1$. In higher dimension, however, $|K_W|$ often says little about W. One instead needs to study $|mK_W|$ for all $m \in \mathbb{Z}$ in order to investigate the geometry of W. This leads to the notion of *Kodaira dimension* $\kappa(W)$, an important birational invariant of W. This is defined to be the maximum of the dimension of the images of W under the maps defined by the linear systems $|mK_W|$ for $m > 0$. It takes values in $\{-\infty, 0, 1, \ldots, \dim X\}$ where the case $-\infty$ corresponds to the situation when $h^0(W, mK_W) = 0$ for every $m > 0$.

Assume $\kappa(W) \geq 0$, that is, $h^0(W, mK_W) \neq 0$ for some $m > 0$. When $m > 0$ is sufficiently divisible, $|mK_W|$ defines a rational fibration $W \dashrightarrow X$ which is called the *Iitaka fibration* of W. This is usually defined up to birational equivalence. The dimension of X is simply the Kodaira dimension $\kappa(W)$. It is often possible to translate questions about W to corresponding questions about X. An old problem is the following:

Conjecture 1.7. *Assume $\kappa(W) \geq 0$. Then there exists $m \in \mathbb{N}$ depending only on* $\dim W$ *such that $|mK_W|$ defines the Iitaka fibration.*

If W is of *general type*, i.e. if $\kappa(W) = \dim W$, then the conjecture is already known by Hacon and McKernan [2006] and Takayama [2006] (also see Hacon, McKernan, and Xu [2013, 2014] for more recent and more general results). In this case we can take m such that $|mK_W|$ defines a birational embedding of W into some projective space. Note that W is birational to its canonical model X, by Birkar, Cascini, Hacon, and McKernan [2010], which is a canonically polarised variety and understanding $|mK_W|$ is the same as understanding $|mK_X|$.

Now assume $0 \le \kappa(W) < \dim W$. The most general known result is that the conjecture is true if we have bounds on certain invariants of the general fibres of the Iitaka fibration, by Birkar and Zhang [2016]. This is done by using a canonical bundle formula for the Iitaka fibration and translating the conjecture into a question on the base of the fibration. Very roughly the main result of Birkar and Zhang [ibid.] says that the conjecture holds if one understands the case $\kappa(W) = 0$. Note that in this case, assuming the minimal model and abundance conjectures, W is birational to a Calabi-Yau variety, and understanding $|mK_W|$ is the same as understanding such systems on the Calabi-Yau variety.

Finally, assume $\kappa(W) = \infty$. Then all the linear systems $|mK_W|$, for $m > 0$, are empty. By the minimal model and abundance conjectures, W is birational to a variety Y admitting a Fano fibration $Y \to Z$. The general fibres of this fibration are Fano varieties. It is then natural to focus on Fano varieties F and study the linear systems $|-mK_F|$, for $m > 0$, in detail. There has been extensive studies of these systems, especially in low dimension, but general higher dimensional results are quite recent; see Birkar [2016a,b].

1.8 Fano varieties, and connection with families, singularities, and termination. Let X be a Fano variety. A difficulty with investigating $|-mK_X|$ is that, unlike the case of varieties of general type, these systems can change dramatically if we change X birationally. On the other hand, a standard inductive technique to study $|-mK_X|$ is to use the elements of $|-mK_X|$ (usually with bad singularities) to create a particular kind of covering family of subvarieties of X and then use induction by restricting to members of this family. A difficulty in this approach is that a member of this family is not necessarily Fano, so it is hard to apply induction, again unlike the case of varieties of general type. Despite these difficulties there has been lots of progress in recent years.

In general there is $m \in \mathbb{N}$ depending only on $\dim X$ such that $|-mK_X|$ is non-empty. Moreover, there is an element of $|-mK_X|$ with good singularities by Birkar [2016a, Theorem 1.1]: this is a special case of *boundedness of complements* (see Theorem 3.3 below). In addition if we put a bound on the singularities of X, that is, if X is ϵ-lc where $\epsilon > 0$, then we can choose m so that $|-mK_X|$ defines a birational embedding of X into some projective space by Birkar [ibid., Theorem 1.2] (see Theorem 3.5 below). In fact one can go further in this case and show that we can choose m so that $-mK_X$ is very ample, hence $|-mK_X|$ defines an embedding of X into some projective space, and that the set of such X form a bounded family by Birkar [2016b, Theorem 1.1]: this is the so-called *BAB conjecture* (see Theorem 3.7 below). These results are proved along with various other results and in conjunction with Shokurov's theory of complements. We will give ample explanations in subsequent sections.

So far we have only mentioned *global Fano* varieties but there are other (relative) Fano varieties. Assume X has good singularities, $f: X \to Z$ is a surjective projective morphism, and $-K_X$ is ample over Z. We call X *Fano over* Z. If Z is a point, then X is a usual Fano variety otherwise in general X is not projective. When $\dim X > \dim Z > 0$, then f is a Fano fibration. Such fibrations appear naturally in birational geometry, and in other contexts, e.g. families and moduli of Fano's.

Now assume f is birational. A special case is a *flipping contraction*, one of the corner stones of the MMP. Existence of flips basically means understanding the linear systems $|-mK_X|$ relatively over Z. Another important special case is when f is the identity morphism in which case we are just looking at the germ of a point on a variety, hence we are doing *singularity theory*. Another connection with singularity theory is that of singularities of $\mathbb{R}$-*linear systems* of divisors on varieties, in general, that is the variety may not be Fano and the divisors may not be related to canonical divisors (see Theorem 4.5 below). This is necessary for the proof of BAB. Therefore, studying Fano varieties in the relative setting naturally overlaps with other important topics in birational and algebraic geometry.

There is also connection with the termination conjecture. It is understood that the termination conjecture is about understanding singularities (see Section 6.5). Moreover, understanding singularities is essentially about understanding Fano varieties in the relative birational case. On the other hand, problems about families of Fano varieties fits well in this theory (see Sections 6.1 and 6.7). It is then no surprise that recent advances on Fano varieties described above is expected to have a profound impact on further developments in birational geometry.

2 Preliminaries

In this section we recall some basic notions. We will try to keep technicalities to a minimum throughout the text. Most of what we need can be found in Kollár and Mori [1998] and Birkar, Cascini, Hacon, and McKernan [2010].

2.1 Contractions. A *contraction* is a projective morphism $f: X \to Z$ of varieties such that $f_*\mathcal{O}_X = \mathcal{O}_Z$. In particular, f is surjective with connected fibres.

2.2 Hyperstandard sets. Let $\mathfrak{R}$ be a subset of $[0,1]$. We define

$$\Phi(\mathfrak{R}) = \left\{1 - \frac{r}{m} \mid r \in \mathfrak{R}, m \in \mathbb{N}\right\}$$

to be the set of *hyperstandard multiplicities* associated to $\mathfrak{R}$. We usually assume $0, 1 \in \mathfrak{R}$ without mention, so $\Phi(\mathfrak{R})$ includes $\Phi(\{0, 1\})$.

2.3 Divisors and resolutions. In algebraic geometry Weil divisors usually have integer coefficients. However, in birational geometry it is standard practice to consider $\mathbb{R}$-divisors. An $\mathbb{R}$-divisor on a normal variety X is of the form $M = \sum a_i M_i$ where M_i are distinct prime divisors and $a_i \in \mathbb{R}$. By $\mu_{M_i} M$ we mean the coefficient a_i. We say M is $\mathbb{R}$-Cartier if M can be written as an $\mathbb{R}$-linear combination of (not necessarily prime) Cartier divisors. For two $\mathbb{R}$-divisors M and N, $M \sim_{\mathbb{R}} N$ means $M - N$ is an $\mathbb{R}$-linear combination of principal Cartier divisors (a principal divisor is the divisor of zeros and poles of a rational function).

If X is equipped with a projective morphism $f : X \to Z$, an $\mathbb{R}$-Cartier divisor M is *nef over* Z if $M \cdot C \geq 0$ for every curve C contracted to a point by f. We say M is *ample over* Z if it is a positive $\mathbb{R}$-linear combination of ample Cartier divisors. We say M is *big over* Z if $M \sim_{\mathbb{R}} A + D$ where A is ample over Z and $D \geq 0$.

A *log resolution* $\phi \colon W \to X$ of (X, M) is a projective birational morphism where W is smooth, and the union of the excpetional locus of ϕ and the birational transform of $\operatorname{Supp} M$ has simple normal crossing singularities.

2.4 Pairs. An important feature of modern birational geometry is that the main objects are pairs rather than varieties. Pairs are much better behaved when it comes to induction and passing from a variety to a birational model.

A *pair* (X, B) consists of a normal variety X and an $\mathbb{R}$-divisor $B \geq 0$ such that $K_X + B$ is $\mathbb{R}$-Cartier. If the coefficients of B are ≤ 1, we say B is a *boundary*.

Let $\phi \colon W \to X$ be a log resolution of (X, B). Let

$$K_W + B_W := \phi^*(K_X + B).$$

The *log discrepancy* of a prime divisor D on W with respect to (X, B) is defines as

$$a(D, X, B) := 1 - \mu_D B_W.$$

We say (X, B) is *lc* (resp. *klt*)(resp. ϵ-*lc*) if every coefficient of B_W is ≤ 1 (resp. < 1)(resp. $\leq 1 - \epsilon$). When $B = 0$ we just say X is lc, etc, instead of $(X, 0)$.

A *non-klt place* of (X, B) is a prime divisor D on birational models of X such that $a(D, X, B) \leq 0$. A *non-klt centre* is the image on X of a non-klt place. When (X, B) is lc, a non-klt centre is also called a *lc centre*.

If we remove the condition $B \geq 0$, the above definitions still make sense but we add *sub* to each notion defined, e.g. instead of lc we say sub-lc, etc.

Example. The simplest kind of pair is a *log smooth* one, that is, a pair (X, B) where X is smooth and Supp B has simple normal crossing singularities. In this case (X, B) being lc (resp. klt) means every coefficient of B is ≤ 1 (resp. < 1).

Example. Let X be the cone over a rational curve of degree n (for a more precise definition see the example following Theorem 3.7). Then X is klt. But if X is the cone over an elliptic curve, then X is lc but not klt.

Example. Let X be a klt surface. Let $\phi\colon W \to X$ be the minimal resolution. The exceptional curves are all smooth rational curves and they intersect in a special way. There is a whose classification of the possible configurations (cf. Kollár and Mori [1998, Section 4]). Once we know the configuration and the self-intersections of the exceptional divisors it is a matter of an easy calculation to determine all the log discrepancies.

2.5 Generalised pairs. These pairs appear mainly when one considers the canonical bundle formula of a fibration, e.g. see case (2) of Section 5.1. A generalised pair is roughly speaking a pair together with a birational polarisation, that is, a nef divisor on some birational model. They play an important role in relation with Conjecture 1.7 by Birkar and Zhang [2016] and most of the results of Birkar [2016a,b]. For the sake of simplicity we will try to avoid using these pairs and their subtle properties as much as possible but for convenience here we recall the definition in the projective case only. For detailed studies of generalised pairs see Birkar and Zhang [2016] and Birkar [2016a].

A projective *generalised (polarised) pair* consists of

- a normal projective variety X',
- an $\mathbb{R}$-divisor $B' \geq 0$ on X',
- a projective birational morphism $\phi\colon X \to X'$ from a normal variety, and
- a nef $\mathbb{R}$-Cartier divisor M on X,

such that $K_{X'} + B' + M'$ is $\mathbb{R}$-Cartier, where $M' := \phi_* M$. We usually refer to the pair by saying $(X', B' + M')$ is a projective generalised pair with data $X \xrightarrow{\phi} X'$ and M. However, we want ϕ and M to be birational data, that is, if we replace X with a higher model, e.g. a resolution, and replace M with its pullback, then we assume the new data defines the same generalised pair.

Now we define generalised singularities. Replacing X we can assume ϕ is a log resolution of (X', B'). We can write

$$K_X + B + M = \phi^*(K_{X'} + B' + M')$$

for some uniquely determined B. We say $(X', B' + M')$ is *generalised lc* (resp. *generalised klt*) if every coefficient of B is ≤ 1 (resp. < 1).

Example. Assume $(X', B' + M')$ is a projective generalised pair with data $X \overset{\phi}{\to} X'$ and M, and assume $M = \phi^* M'$. Then $(X', B' + M')$ is generalised lc (resp. generalised klt) iff (X', B') is lc (resp. klt). In other words, in this case M' does not contribute to singularities.

Example. Let $X' = \mathbb{P}^2$, and $\phi\colon X \to X'$ be the blowup of a closed point $x' \in X'$. Assume $H \subset$ is a hyperplane. Let $M = 3\phi^* H - tE$ where E is the exceptional divisor of ϕ and $t \in [0, 3]$ is a real number. Then letting $B' = 0$, $(X', B' + M')$ is a projective generalised pair with data $X \overset{\phi}{\to} X'$ and M. Note that $M' = 0$. Now we can determine B in the formula above. Calculating intersection numbers we find $B = (t-1)E$. Therefore, $(X', B' + M')$ is generalised lc (resp. generalised klt) iff $t \leq 2$ (resp. $t < 2$).

3 Fano varieties

3.1 Facets of Fano varieties. Grothendieck insisted on studying varieties (and schemes) in a relative setting. This philosophy has been very successfully implemented in birational geometry. This is particularly interesting in the case of Fano varieties, or we should say relative Fano varieties.

Let (X, B) be a klt pair and $f\colon X \to Z$ be a surjective projective morphism, and assume $-(K_X + B)$ is ample over Z. We then say (X, B) is *Fano over Z*. This relative notion unifies various classes of objects of central importance. There are three distinct cases.

- *Global case:* this is when Z is just a point, hence (X, B) is a *Fano pair* in the usual sense.
- *Fibration case:* this is when $\dim X > \dim Z > 0$, that is, f is a genuine fibration and its general fibres are global Fano pairs.
- *Birational case:* this is when f is birational. There are several important subcases here. If f is extremal and contracts one divisor, then f is a *divisorial contraction*. If f is extremal and contracts some subvariety but not a divisor, then f is a *flipping contraction*. If f is an isomorphism, then (X, B) is just the germ of a klt singularity.

3.2 Complements and anti-pluri-canonical systems. Assume (X, B) is an lc pair equipped with a projective morphism $X \to Z$. The *theory of complements* is essentially the study of the systems $|-n(K_X+B)|$ where $n \in \mathbb{N}$, in a relative sense over Z. Obviously this is interesting only when some of these systems are not non-empty, e.g. Fano case. The theory was introduced by Shokurov [1993]. The theory was further developed by Shokurov [2000], Y. G. Prokhorov and Shokurov [2001, 2009], and Birkar [2016a,b].

A *strong n-complement* of K_X+B over a point $z \in Z$ is of the form K_X+B^+ where over some neighbourhood of z we have:

- (X, B^+) is lc,
- $n(K_X+B^+) \sim 0$, and
- $B^+ \geq B$.

From the definition we get

$$-n(K_X+B) \sim nB^+ - nB \geq 0$$

over some neighbourhood of z which in particular means the linear system $|-n(K_X+B)|$ is not empty over z, and that it contains a "nice" element. An n-complement (see Birkar [2016a]) is defined similarly but it is more complicated, so for simplicity we avoid using it. However, if $B=0$, a complement and a strong complement are the same thing.

Theorem 3.3 (Birkar [ibid., Theorems 1.7, 1.8, 1.9])**.** *Let d be a natural number and $\mathfrak{R} \subset [0,1]$ be a finite set of rational numbers. Then there exists a natural number n depending only on d and $\mathfrak{R}$ satisfying the following. Assume (X,B) is a pair and $X \to Z$ a contraction such that*

- *(X,B) is lc of dimension d,*
- *the coefficients of B are in $\Phi(\mathfrak{R})$,*
- *X is Fano type over Z, and*
- *$-(K_X+B)$ is nef over Z.*

Then for any point $z \in Z$, there is a strong n-complement K_X+B^+ of K_X+B over z. Moreover, the complement is also an mn-complement for any $m \in \mathbb{N}$.

Here X of *Fano type* over Z means (X,G) is Fano over Z for some G. The theorem was conjectured by Shokurov [2000, Conjecture 1.3], it was proved in dimension 2 by Shokurov [ibid., Theorem 1.4], (see also Y. G. Prokhorov and Shokurov [2009, Corollary 1.8], and Shokurov [1993] for some cases). Y. G. Prokhorov and Shokurov [2001, 2009] proved various inductive statements regarding complements including some unconditional cases in dimension 3.

Example. When $X \to Z$ is toric morphism and $B = 0$ we can take $n = 1$ and B^+ to be the sum of the torus-invariant divisors.

Remark. Assume Z is a point. Assume for simplicity that $B = 0$ and that $-K_X$ is ample, that is, X is a usual Fano variety. When X is a smooth 3-fold, Šokurov [1979] proved that $|-K_X|$ contains a smooth K3 surface. In particular, K_X has a 1-complement. This is probably where the higher dimensional theory of complements originates.

Remark. Assume $X \to Z$ is birational. Assume again for simplicity that $B = 0$ and that $-K_X$ is ample over Z. When $X \to Z$ is a flipping contraction contracting one smooth rational curve only, Mori (cf. Kollár and Mori [1992, Theorem 1.7]) showed that there always exists a 1-complement over each $z \in Z$ but in the analytic sense, i.e. it exists over an analytic neighbourhood of z. This is used in Mori's proof of existence of 3-fold flips, see Mori [1988].

Remark. Assume $X \to Z$ is an isomorphism, so we are looking at the germ of a klt singularity (X, B) around a point $x \in X$. For simplicity again assume $B = 0$. In general the Cartier index of K_X is not bounded even in dimension 2. The point of complement theory in this case is that the n-complement $K_X + B^+$ has Cartier index n which is bounded.

Remark. When X is a 3-fold with terminal singularities, $-K_X$ is ample over Z, and $B = 0$, the *general elephant conjecture* of Reid asks whether a general element of the linear system $|-K_X|$, relatively over Z, has canonical singularities. This is true in various cases, e.g. when X is Gorenstein and Z is a point by Reid [n.d.], or when $X \to Z$ is identity by Reid [1987].

Example. Lets look at the particular case of surfaces in the local case. Assume X is a surface, $X \to Z$ is the identity, and $B = 0$. If $x \in X$ is smooth, then K_X is a 1-complement of itself, that is, we can take $B^+ = 0$. In the singular case, from classification of the possible singularities one gets the following, by Shokurov [2000, p. 5.2.3]:

$$\text{if } x \in X \text{ is a type} \begin{cases} A \text{ singularity, then } K_X \text{ has a 1-complement.} \\ D \text{ singularity, then } K_X \text{ has a 2-complement.} \\ E_6 \text{ singularity, then } K_X \text{ has a 3-complement.} \\ E_7 \text{ singularity, then } K_X \text{ has a 4-complement.} \\ E_8 \text{ singularity, then } K_X \text{ has a 6-complement.} \end{cases}$$

3.4 Effective birationality. Let X be a Fano variety. Theorem 3.3 says that $|-mK_X|$ is non-empty containing a nice element for some $m > 0$ depending only on $\dim X$. If we bound the singularities of X, we then have a much stronger statement.

Theorem 3.5 (Birkar [2016a, Theorem 1.2]). *Let d be a natural number and $\epsilon > 0$ a real number. Then there is a natural number m depending only on d and ϵ such that if X is any ϵ-lc Fano variety of dimension d, then $|-mK_X|$ defines a birational map.*

Note that m indeed depends on d as well as ϵ because the theorem implies the volume $\mathrm{vol}(-K_X)$ is bounded from below by $\frac{1}{m^d}$. Without the ϵ-lc assumption, $\mathrm{vol}(-K_X)$ can get arbitrarily small or large; see Hacon, McKernan, and Xu [2014, Example 2.1.1]. In dimension 2, the theorem is a consequence of BAB proved by V. Alexeev [1994], and in dimension 3, special cases are proved by Jiang [2015] using different methods. Cascini and McKernan have independently proved the theorem for canonical singularities, that is when $\epsilon = 1$, using quite different methods.

It is worth mentioning that the theorem also holds in the relative setting. It follows immediately from the global case stated above.

3.6 Boundedness of Fano varieties: BAB. It is possible to strengthen Theorem 3.5 so that $|-mK_X|$ defines an actual embedding. This follows from the next result.

Theorem 3.7 (Birkar [2016b, Theorem 1.1]). *Let d be a natural number and ϵ a positive real number. Then the projective varieties X such that*

- *(X, B) is ϵ-lc of dimension d for some boundary B, and*
- *$-(K_X + B)$ is nef and big,*

form a bounded family.

This was known as the Borisov-Alexeev-Borisov or BAB conjecture. Various special cases of it was considered by many people. It was known in the following cases (by taking $B = 0$): surfaces by V. Alexeev [1994], toric varieties by A. A. Borisov and L. A. Borisov [1992], Fano 3-folds with terminal singularities and Picard number one by Kawamata [1992b], Fano 3-folds with canonical singularities by Kollár, Miyaoka, Mori, and Takagi [2000], smooth Fano varieties by Kollár, Miyaoka, and Mori [1992], spherical Fano varieties by V. A. Alexeev and Brion [2004], Fano 3-folds with fixed Cartier index of K_X by A. Borisov [2001], and more generally, Fano varieties of given dimension with fixed Cartier index of K_X by Hacon, McKernan, and Xu [2014]; in a given dimension, the Fano varieties X equipped with a boundary Δ such that $K_X + \Delta \equiv 0$, (X, Δ) is ϵ-lc, and such that the coefficients of Δ belong to a DCC set, by Hacon, McKernan, and Xu [ibid.] (also see Hacon and Xu [2015] and Birkar [2016a, Theorem 1.4]).

Example. Now we look at an example of a non-bounded family of singular Fano surfaces. For $n \geq 2$ consider

$$\begin{array}{ccc} E & \subset W_n & \xrightarrow{f} X_n \\ & \downarrow & \\ & \mathbb{P}^1 & \end{array}$$

where X_n is the cone over a rational curve of deg n, f is blowup of the vertex, and E is the exceptional curve. In other words, W_n is the projective bundle of $\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-n)$, E is the section given by the summand $\mathcal{O}_{\mathbb{P}^1}(-n)$, and X_n is obtained from W_n by contracting E. Then an easy calculation, using $E^2 = -n$, shows that

$$K_{W_n} + \frac{n-2}{n} E = f^* K_{X_n},$$

hence X_n is a $\frac{2}{n}$-lc Fano variety with one singular point (the larger is n, the deeper is the singularity). In particular, since the set of numbers $\{\frac{n-2}{n} \mid n \in \mathbb{N}\}$ is not finite, the set $\{X_n \mid n \in \mathbb{N}\}$ is not a bounded family. This example explains the role of the number ϵ in Theorem 3.7.

Example. In this example we sketch the proof of Theorem 3.7 in dimension two following V. Alexeev and Mori [2004]. For simplicity assume $B = 0$ and that $-K_X$ is ample. There is $\Delta \geq 0$ such that (X, Δ) is ϵ-lc and $K_X + \Delta \sim_{\mathbb{R}} 0$. Let $\phi\colon W \to X$ be the minimal resolution and let $K_W + \Delta_W$ be the pullback of $K_X + \Delta$. Since (X, Δ) is klt, the exceptional divisors of ϕ are all smooth rational curves. Moreover, by basic properties of minimal resolutions, $\Delta_W \geq 0$. In particular, (W, Δ_W) is an ϵ-lc pair. Now a simple calculation of intersection numbers shows that $-E^2 \leq l$ for every exceptional curve of ϕ where $l \in \mathbb{N}$ depends only on ϵ. If the number of exceptional curves of ϕ is bounded, then the Cartier index of $-K_X$ is bounded which in turn implies $-nK_X$ is very ample for some bounded n. In particular, this holds if the Picard number of W is bounded from above. If in addition $\mathrm{vol}(-K_X)$ is bounded, then X belongs to a bounded family. Note that $\mathrm{vol}(-K_X) = \mathrm{vol}(-K_W)$.

Running an MMP on K_W we get a morphism $W \to V$ where V is either $\mathbb{P}^2$ or a rational ruled surface (like W_n in the previous example), and the morphism is a sequence of blowups at smooth points. Let Δ_V be the pushdown of Δ_W. Then (V, Δ_V) is ϵ-lc and $K_V + \Delta_V \sim_{\mathbb{R}} 0$. It is easy to show that there are finitely many possibilities for V. In particular, from $\mathrm{vol}(-K_W) \leq \mathrm{vol}(-K_V)$, we deduce that $\mathrm{vol}(-K_X) = \mathrm{vol}(-K_W)$ is bounded from above. Thus it is enough to prove that the number of blowups in $W \to V$ is bounded. This number can be bounded by an elementary analysis of possible intersection numbers in the sequence (see V. Alexeev and Mori [ibid., Section 1] for more details).

3.8 Birational automorphism groups. An interesting consequence of Theorem 3.7 concerns the Jordan property of birational automorphism groups of rationally connected varieties. Y. Prokhorov and C. Shramov [2016, Theorem 1.8] proved the next result assuming Theorem 3.7.

Corollary 3.9 (Birkar [2016b, Corollary 1.3]). *Let d be a natural number. Then there is a natural number h depending only on d satisfying the following. Let X be a rationally connected variety of dimension d over k. Then for any finite subgroup G of the birational automorphism group* $\mathrm{Bir}(X)$*, there is a normal abelian subgroup H of G of index at most h. In particular,* $\mathrm{Bir}(X)$ *is Jordan.*

Here X *rationally connected* means that every two general closed points can be joined by a rational curve. If we take $X = \mathbb{P}^d$ in the corollary, then we deduce that the Cremona group $\mathrm{Cr}_d := \mathrm{Bir}(\mathbb{P}^d)$ is Jordan, answering a question of Serre [2009, p. 6.1].

4 Singularities of linear systems

4.1 Lc thresholds of $\mathbb{R}$-linear systems. Let (X, B) be a pair. The *log canonical threshold* (lc threshold for short) of an $\mathbb{R}$-Cartier $\mathbb{R}$-divisor $L \geq 0$ with respect to (X, B) is defined as

$$\mathrm{lct}(X, B, L) := \sup\{t \mid (X, B + tL) \text{ is lc}\}.$$

It is a way of measuring the singularities of L taking into account the singularities of (X, B) as well.

Now let A be an $\mathbb{R}$-Cartier $\mathbb{R}$-divisor. The $\mathbb{R}$-linear system of A is

$$|A|_{\mathbb{R}} = \{L \geq 0 \mid L \sim_{\mathbb{R}} A\}.$$

We then define the *lc threshold* of $|A|_{\mathbb{R}}$ with respect to (X, B) (also called global lc threshold or α-invariant) as

$$\mathrm{lct}(X, B, |A|_{\mathbb{R}}) := \inf\{\mathrm{lct}(X, B, L) \mid L \in |A|_{\mathbb{R}}\}$$

which coincides with

$$\sup\{t \mid (X, B + tL) \text{ is lc for every } L \in |A|_{\mathbb{R}}\}.$$

This is an asymptotic invariant, so not surprisingly it is hard to compute in specific cases and study in general.

Due to connections with the notion of stability and existence of Kähler-Einstein metrics, lc thresholds of $\mathbb{R}$-linear systems have attracted a lot of attention, particularly, when A is ample. An important special case is when X is Fano and $A = -K_X$ in which case many examples have been calculated, e.g. see Cheltsov and K. A. Shramov [2008].

Example. If $X = \mathbb{P}^d$, $B = 0$, and $A = -K_X$, then

$$\operatorname{lct}(X, B, |A|_{\mathbb{R}}) = \frac{1}{d+1}.$$

On the other hand, if $X \subset \mathbb{P}^d$ is a smooth hypersurface of degree $r \leq d$, $B = 0$, and $A = -K_X$, then

$$\operatorname{lct}(X, B, |A|_{\mathbb{R}}) = \frac{1}{d+1-r}$$

by Cheltsov and K. A. Shramov [ibid., Example 1.3].

Another reason for studying the above threshold is connection with boundedness of Fano varieties. Indeed it plays a central role in the proof of Theorem 3.7.

Theorem 4.2 (Birkar [2016b, Theorem 1.4]). *Let d be a natural number and ϵ a positive real number. Then there is a positive real number t depending only on d, ϵ satisfying the following. Assume*

- *(X, B) is a projective ϵ-lc pair of dimension d, and*
- *$A := -(K_X + B)$ is nef and big.*

Then

$$\operatorname{lct}(X, B, |A|_{\mathbb{R}}) \geq t.$$

This was conjectured by Ambro [2016] who proved it in the toric case. It can be derived from Theorem 3.7 but in reality it is proved before Theorem 3.7 (see next section). Jiang [2015, 2014] proved it in dimension two.

The lc threshold of an $\mathbb{R}$-linear system $|A|_{\mathbb{R}}$ is defined as an infimum of usual lc thresholds. Tian [1990, Question 1] asked whether the infimum is a minimum when $A = -K_X$ and X is Fano. The question was reformulated and generalised to Fano pairs in Cheltsov and K. A. Shramov [2008, Conjecture 1.12]. The next result gives a positive answer when the lc threshold is at most 1.

Theorem 4.3 (Birkar [2016b, Theorem 1.5]). *Let (X, B) be a projective klt pair such that $A := -(K_X + B)$ is nef and big. Assume that* $\operatorname{lct}(X, B, |A|_{\mathbb{R}}) \leq 1$. *Then there is $0 \leq D \sim_{\mathbb{R}} A$ such that*

$$\operatorname{lct}(X, B, |A|_{\mathbb{R}}) = \operatorname{lct}(X, B, D).$$

Moreover, if B is a $\mathbb{Q}$-boundary, then we can choose $D \sim_{\mathbb{Q}} A$, hence in particular, the lc threshold is a rational number.

Shokurov has an unpublished proof of the theorem in dimension two.

4.4 Lc thresholds of $\mathbb{R}$-linear systems with bounded degree. Next we treat lc thresholds associated with divisors on varieties, in a general setting. To obtain any useful result, one needs to impose certain boundedness conditions on the invariants of the divisor and the variety.

Theorem 4.5 (Birkar [2016b, Theorem 1.6]). *Let d, r be natural numbers and ϵ a positive real number. Then there is a positive real number t depending only on d, r, ϵ satisfying the following. Assume*

- *(X, B) is a projective ϵ-lc pair of dimension d,*
- *A is a very ample divisor on X with $A^d \le r$, and*
- *$A - B$ is ample.*

Then

$$\operatorname{lct}(X, B, |A|_{\mathbb{R}}) \ge t.$$

This is one of the main ingredients of the proof of Theorem 4.2 but it is also interesting on its own. We explain briefly some of the assumptions of the theorem. The condition $A^d \le r$ means that X belongs to a bounded family of varieties, actually, if we choose A general in its linear system, then (X, A) belongs to a bounded family of pairs. We can use the divisor A to measure how "large" other divisors are on X. Indeed, the ampleness of $A - B$ roughly speaking says that the "degree" of B is bounded from above, that is,

$$\deg_A B := A^{d-1}B < A^d \le r.$$

Without such boundedness assumptions, one would not find a positive lower bound for the lc threshold as the next example shows.

Example. Assume $(X = \mathbb{P}^2, B)$ is ϵ-lc and $S \subset X$ is a line. Let $L = A = lS$ where $l \in \mathbb{N}$. Then the multiplicity of L at any closed point $x \in L$ is l, hence the lc threshold $\operatorname{lct}(L, X, B) \le \frac{1}{l}$. Thus the larger is l, the smaller is the threshold. Next we illustrate how the threshold depends on the degree of B. Let T be another line and x be the intersection point $S \cap T$. Let $X_1 \to X$ be the blowup at x, and let x_1 be the intersection of the exceptional divisor E_1 and the birational transform $S^\sim$. Let $X_2 \to X_1$ be the blowup at x_1, and let x_2 be the intersection of the new exceptional divisor E_2 and $S^\sim$. At each step we blowup the intersection point of $S^\sim$ and the newest exceptional divisor.

Put $W := X_r$. Then the exceptional locus of $\phi\colon W \to X$ consists of a chain of curves all of which are -2-curves except one which is a -1-curve. Then $-K_W$ is nef over X, in fact, it is semi-ample over X. Thus there is $0 \le B_W \sim_{\mathbb{R}} \alpha\phi^* H - K_W$ for some $\alpha > 0$ such that (W, B_W) is $\frac{1}{2}$-lc and $K_W + B_W \sim_{\mathbb{R}} 0/X$. Now let B be the pushdown of B_W.

Then (X, B) is $\frac{1}{2}$-lc. Now let $L = S + T$. Then the coefficient of E_r in $\phi^* L$ is $r + 1$, hence

$$\operatorname{lct}(L, X, B) = \operatorname{lct}(\phi^* L, W, B_W) \leq \frac{1}{r+1}.$$

Thus there is no lower bound on the lc threshold if r is arbitrarily large. This does not contradict Theorem 4.5 because when $r \gg 0$, the degree $\deg_A B \gg 0$ and $A - B$ cannot be ample (here $A = lS$ with l fixed).

5 Brief sketch of proofs of main results

In this section we sketch some of the ideas of the proofs of Theorem 3.3, Theorem 3.5, Theorem 4.5 and Theorem 3.7. We try to remove technicalities as much as possible but this comes at the expense of being imprecise in various places and not elaborating on many of the new ideas.

5.1 Sketch of proof of boundedness of complements. (Theorem 3.3) For simplicity we look at the global case, that is, when Z is a point. Pick a sufficiently small $\epsilon \in (0, 1)$. Let $Y \to X$ be the birational morphism which extracts all the prime divisors with log discrepancy smaller than ϵ. Let $K_Y + B_Y$ be the pullback of $K_X + B$. Define Θ_Y to be the same as B_Y except that we replace each coefficient in $(1-\epsilon, 1)$ with 1. Run an MMP on $-(K_Y + \Theta_Y)$ and let Y' be the resulting model and $\Theta_{Y'}$ be the pushdown of Θ_Y. We can run such MMP because Y turns out to be of Fano type, so we can run MMP on any divisor on Y.

As a consequence of local and global ACC of Hacon, McKernan, and Xu [2014, Theorems 1.1 and 1.5] (in practice we need their generalisations to generalised pairs, see Birkar and Zhang [2016, Theorems 1.5 and 1.6]), we can show that the MMP does not contract any component of $\lfloor \Theta_Y \rfloor$, $(Y', \Theta_{Y'})$ is lc, and $-(K_{Y'} + \Theta_{Y'})$ is nef. It is enough to construct a bounded complement for $K_{Y'} + \Theta_{Y'}$. Replacing (X, B) with $(Y', \Theta_{Y'})$ and applying further reductions, we can reduce the problem to one of the following cases:

1. B has a component S with coefficient 1 and $-(K_X + B)$ is nef and big, or
2. $K_X + B \equiv 0$ along a fibration $f : X \to T$, or
3. (X, B) is *exceptional*.

Here exceptional means that for any choice of $0 \leq P \sim_{\mathbb{R}} -(K_X + B)$ the pair $(X, B + P)$ is klt. These cases require very different inductive treatment.

Case (1): First apply *divisorial adjunction* to define $K_S + B_S = (K_X + B)|_S$. Further modification of the setting allows us to ensure that S is Fano type. Moreover, the coefficients of B_S happen to be in a set $\Phi(\mathfrak{S})$ for some fixed finite set $\mathfrak{S}$. By induction on dimension $K_S + B_S$ has a strong n-complement for some bounded n. The idea then is to lift the complement to X using vanishing theorems. In the simplest case when (X, B) is log smooth and $B = S$, we look at the exact sequence

$$H^0(-n(K_X + B)) \to H^0(-n(K_X + B)|_S) \to H^1(-n(K_X + B) - S) = 0$$

where the vanishing follows from Kawamata-Viehweg vanishing theorem noting that

$$-n(K_X + B) - S = K_X - n(K_X + B) - (K_X + B) = K_X - (n+1)(K_X + B)$$

Since $K_S + B_S$ has a strong n-complement, the middle space in the above sequence is non-trivial which implies the left hand side is also non-trivial by lifting the section corresponding to the complement. One then argues that the lifted section gives a strong n-complement for $K_X + B$.

Case (2): Apply the *canonical bundle formula* (also called adjunction for fibre spaces, derived from Kawamata [1998]) to write

$$K_X + B \sim_{\mathbb{R}} f^*(K_T + B_T + M_T)$$

where B_T is the *discriminant divisor* and M_T is the *moduli divisor*. It turns out that the coefficients of B_T happen to be in a set $\Phi(\mathfrak{S})$ for some fixed finite set $\mathfrak{S}$, and that pM_T is integral for some bounded number $p \in \mathbb{N}$. Now we want to find a complement for $K_T + B_T + M_T$ and pull it back to X. There is a serious issue here: $(T, B_T + M_T)$ is not a pair in the usual sense but it is a generalised pair. Thus we actually need to prove Theorem 3.3 (at least in the global case) in the more general setting of generalised pairs. This makes life a lot more difficult but fortunately everything turns out to work. Once we have a bounded complement for $K_T + B_T + M_T$ it is straightforward to derive a bounded complement for $K_X + B$.

Case (3): In this case we use effective birationality. Perhaps after decreasing ϵ, the exceptionality condition implies that (X, B) is ϵ-lc. For simplicity assume $B = 0$ and that X is a Fano variety. Also assume we already have Theorem 3.5. Then there is a bounded number $m \in \mathbb{N}$ such that $|-mK_X|$ defines a birational map. Pick $M \in |-mK_X|$ and let $B^+ = \frac{1}{m}M$. Since X is exceptional, (X, B^+) is automatically klt, hence $K_X + B^+$ is a strong m-complement. Although this gives some ideas of how one may get a bounded complement but in practice we cannot give a complete proof of Theorem 3.5 before proving Theorem 3.3. The two theorems are actually proved together. See Birkar [2016a, Sections 6 and 7] for more details.

5.2 Sketch of proof of effective birationality. (Theorem 3.5) Let $m \in \mathbb{N}$ be the smallest number such that $|-mK_X|$ defines a birational map, and let $n \in \mathbb{N}$ be a number such that $\mathrm{vol}(-nK_X) > (2d)^d$. Initially we take n to be the smallest such number but we will modify it during the proof. We want to show that m is bounded from above. The idea is first to show that $\frac{m}{n}$ is bounded from above, and then at the end show that m is bounded.

Applying a standard elementary technique we can create a covering family $\mathcal{G}$ of subvarieties of X such that if $x, y \in X$ are any pair of general closed points, then there is $0 \le \Delta \sim_{\mathbb{Q}} -(n+1)K_X$ and $G \in \mathcal{G}$ such that (X, Δ) is lc at x with the unique non-klt centre G, and (X, Δ) is not klt at y.

Assume $\dim G = 0$ for all G. Then $G = \{x\}$ is an isolated non-klt centre. Using multiplier ideals and vanishing theorems we can lift sections from G and show that $|-nK_X|$ defines a birational map after replacing n with a bounded multiple, hence in particular $\frac{m}{n}$ is bounded from above in this case.

Now lets assume all G have positive dimension. If $\mathrm{vol}(-mK_X|_G)$ is large, then again using some elementary arguments, we can create a new non-klt centre G' containing x but with $\dim G' < \dim G$. Thus we can replace G with G' and apply induction on dimension of G. We can then assume $\mathrm{vol}(-mK_X|_G)$ is bounded from above.

Similar to the previous paragraph, we can cut G and decrease its dimension if $\mathrm{vol}(-nK_X|_G)$ is bounded from below. Showing this lower boundedness is the hard part. A key point here is that although G is not necessarily a divisor and although the singularities of (X, Δ) away from x maybe quite bad but still there is a kind of adjunction formula, that is, if F is the normalisation of G, then we can write

$$(K_X + \Delta)|_F \sim_{\mathbb{R}} K_F + \Theta_F + P_F$$

where Θ_F is a boundary divisor with coefficients in a fixed DCC set Ψ depending only on d, and P_F is pseudo-effective. Replacing n with $2n$ and adding to Δ we can easily make P_F big and effective.

Now we would ideally want to apply induction on d but the difficulty is that F may not be Fano, in fact, it can be any type of variety. Another issue is that the singularities of $(F, \Theta_F + P_F)$ can be pretty bad. To overcome these difficulties we use the fact that $\mathrm{vol}(-mK_X|_G)$ is bounded from above. From this boundedness one can deduce that there is a bounded projective log smooth pair $(\overline{F}, \Sigma_{\overline{F}})$ and a birational map $\overline{F} \dashrightarrow F$ such that $\Sigma_{\overline{F}}$ is reduced containing the exceptional divisor of $\overline{F} \dashrightarrow F$ and the support of the birational transform of Θ_F (and other relevant divisors).

Surprisingly, the worse the singularities of $(F, \Theta_F + P_F)$ the better because we can then produce divisors on $\overline{F}$ with bounded "degree" but with arbitrarily small lc thresholds which would contradict a baby version of Theorem 4.5. Indeed assume $(F, \Theta_F + P_F)$ is not klt. A careful study of the above adjunction formula allows to write $K_F + \Lambda_F := K_X|_F$ where $\Lambda_F \le \Theta_F$ and (F, Λ_F) is sub-ϵ-lc. Put $I_F = \Theta_F + P_F - \Lambda_F$. Then

$$I_F = K_F + \Theta_F + P_F - K_F - \Lambda_F \sim_{\mathbb{R}} (K_X + \Delta)|_F - K_X|_F = \Delta|_F \sim_{\mathbb{R}} -(n+1)K_X|_F.$$

Moreover, $K_F + \Lambda_F + I_F$ is ample.

Let $\phi\colon F' \to F$ and $\psi\colon F' \to \overline{F}$ be a common resolution. Pull back $K_F+\Lambda_F+I_F$ to F' and then push it down to $\overline{F}$ and write it as $K_{\overline{F}}+\Lambda_{\overline{F}}+I_{\overline{F}}$. Then the above ampleness gives

$$\phi^*(K_F+\Lambda_F+I_F) \le \psi^*(K_{\overline{F}}+\Lambda_{\overline{F}}+I_{\overline{F}})$$

which implies that $(\overline{F}, \Lambda_{\overline{F}}+I_{\overline{F}})$ is not sub-klt. From this one deduces that $(\overline{F}, \Gamma_{\overline{F}}+I_{\overline{F}})$ is not klt where $\Gamma_{\overline{F}} = (1-\epsilon)\Sigma_{\overline{F}}$. Finally, one argues that the degree of $I_{\overline{F}}$ gets arbitrarily small if $\mathrm{vol}(-nK_X|_G)$ gets arbitrarily small, and this contradicts an easy case of Theorem 4.5.

If singularities of $(F, \Theta_F + P_F)$ are good, then we again face some serious difficulties. Very roughly, in this case, we lift sections from F to X and use this section to modify Δ so that $(F, \Theta_F + P_F)$ has bad singularities, hence we reduce the problem to the above arguments. This shows $\frac{m}{n}$ is bounded.

Finally, to we still need to bound m. This can be done by arguing that $\mathrm{vol}(-mK_X)$ is bounded from above and use this to show X is birationally bounded, and then work on the bounded model. See Birkar [2016a, Section 4] for more details.

5.3 Sketch of proof of boundedness of lc thresholds. (Theorem 4.5) Pick $0 \le N \sim_{\mathbb{R}} A$. Let s be the largest number such that $(X, B+sN)$ is ϵ'-lc where $\epsilon' = \frac{\epsilon}{2}$. It is enough to show s is bounded from below. There is a prime divisor T on birational models of X with log discrepancy $a(T, X, \Delta) = \epsilon'$ where $\Delta := B + sN$. It is enough to show that the multiplicity of T in ϕ^*N is bounded on some resolution $\phi\colon V \to X$ on which T is a divisor. We can assume the image of T on X is a closed point x otherwise we can cut by hyperplane sections and apply induction on dimension.

There is a birational morphism $Y \to X$ from a normal projective variety which contracts exactly T. A key ingredient here is provided by the theory of complements: using the fact that $-(K_Y+T)$ is ample over X, we can find Λ_Y such that (Y, Λ_Y) is lc near T and $n(K_Y+\Lambda_Y) \sim 0/X$ for some bounded number $n \in \mathbb{N}$. One can think of $K_Y+\Lambda_Y$ as a local-global type of complement. The crucial point is that if Λ is the pushdown of Λ_Y, then we can make sure degree of Λ is bounded from above, that is, after replacing A we can assume $A - \Lambda$ is ample. By construction, the log discrepancy $a(T, X, \Lambda) = 0$ and $(X, \operatorname{Supp}\Lambda)$ is bounded.

Next using resolution of singularities we can modify the setting and then assume that (X, Λ) is log smooth and Λ is reduced. The advantage of having Λ is that now T can be obtained by a sequence of blowups which is toroidal with respect to (X, Λ). That is, in every step we blowup the centre of T which happens to be a stratum of (X, Λ); a stratum is just a component of the intersection of some of the components of Λ. The first step is just the blowup of x. One argues that it is enough to bound the number of these blowups.

By the previous paragraph, we can discard any component of Λ not passing through x, hence assume $\Lambda = S_1 + \cdots + S_d$ where S_i are irreducible components. On the other hand, a careful analysis of $Y \to X$ allows us to further modify the situation so that $\operatorname{Supp} \Delta$ does not contain any stratum of (X, Λ) apart from x. This is one of the difficult steps of the whole proof.

Since (X, Λ) is log smooth and bounded, we can find a surjective finite morphism $X \to \mathbb{P}^d$ which maps x to the origin $z = (0 : \cdots : 0 : 1)$ and maps S_i on H_i where $H_1, \ldots, H_d$ are the coordinate hyperplanes passing through z. Since $\operatorname{Supp} \Delta$ does not contain any stratum of (X, Λ) apart from x, it is not hard to reduce the problem to a similar problem on $\mathbb{P}^d$. From now on we assume $X = \mathbb{P}^d$ and that S_i are the coordinate hyperplanes. The point of this reduction is that now (X, Λ) is not only toroidal but actually toric, and $-(K_X + \Lambda)$ is very ample. In particular, replacing Δ with $t\Delta + (1-t)\Lambda$ for some sufficiently small $t > 0$ (and replacing ϵ' accordingly), we can make $K_X + \Delta$ anti-ample. Next by adding to Δ we can assume $K_X + \Delta$ is numerically trivial.

Let $W \to X$ be the sequence of blowups which obtains T as above. Since the blowups are toric, W is a toric variety. If $Y \to X$ is the birational morphism contracting T only, as before, then Y is also a toric variety. Moreover, if $K_Y + \Delta_Y$ is the pullback of $K_X + \Delta$, then (Y, Δ_Y) is ϵ'-lc and $K_Y + \Delta_Y$ is numerically trivial. Now running MMP on $-K_Y$ and using base point freeness gives another toric variety Y' which is Fano and ϵ'-lc. By the toric version of BAB proved by A. A. Borisov and L. A. Borisov [1992], Y' belongs to a bounded family. From this we can produce a klt strong m-complement $K_{Y'} + \Omega_{Y'}$ for some bounded $m \in \mathbb{N}$ which induces a klt strong m-complement $K_Y + \Omega_Y$ which in turn gives a klt strong m-complement $K_X + \Omega$.

Finally Ω belongs to a bounded family as its coefficients are in a fixed finite set and its degree is bounded. This implies that $(X, \Omega + u\Lambda)$ is klt for some $u > 0$ bounded from below. Now an easy calculation shows that the multiplicity of T in the pullback of Λ on W is bounded from above which in turn implies the number of blowups in $W \to X$ is bounded as required.

5.4 Sketch of proof of BAB. (Theorem 3.7) First applying Hacon and Xu [2015, Theorem 1, 3] it is enough to show that K_X has a klt strong m-complement for some bounded number $m \in \mathbb{N}$. Running an MMP on $-K_X$ and replacing X with the resulting model we can assume $B = 0$. By Theorem 3.3, we know that we have an lc strong n-complement $K_X + B^+$. If X is exceptional, then the complement is klt, so we are done in this case. To treat the general case the idea is to modify the complement $K_X + B^+$ into a klt one. We will do this using birational boundedness.

We need to show $\operatorname{vol}(-K_X)$ is bounded from above. This can be proved using arguments similar to the proof of the effective birationality theorem. Once we have this bound, we can show that (X, B^+) is log birationally bounded, that is, there exist a bounded log smooth projective pair $(\overline{X}, \Sigma_{\overline{X}})$ and a birational map $\overline{X} \dashrightarrow X$ such that $\Sigma_{\overline{X}}$ contains the exceptional divisors of $\overline{X} \dashrightarrow X$ and the support of the birational transform of B^+.

Next we pull back $K_X + B^+$ to a high resolution of X and push it down to $\overline{X}$ and denote it by $K_{\overline{X}} + B^+_{\overline{X}}$. Then $(\overline{X}, B^+_{\overline{X}})$ is sub-lc and $n(K_{\overline{X}} + B^+_{\overline{X}}) \sim 0$. Now support of $B^+_{\overline{X}}$ is contained in $\Sigma_{\overline{X}}$ so we can use the boundedness of $(\overline{X}, \Sigma_{\overline{X}})$ to perturb the coefficients of $B^+_{\overline{X}}$. More precisely, perhaps after replacing n, there is $\Delta_{\overline{X}} \sim_{\mathbb{Q}} B^+_{\overline{X}}$ such that $(\overline{X}, \Delta_{\overline{X}})$ is sub-klt and $n(K_{\overline{X}} + \Delta_{\overline{X}}) \sim 0$. Pulling $K_{\overline{X}} + \Delta_{\overline{X}}$ back to X and denoting it by $K_X + \Delta$ we get a sub-klt (X, Δ) with $n(K_X + \Delta) \sim 0$.

Now a serious issue here is that Δ is not necessarily effective. In fact it is by no means clear that its coefficients are even bounded from below. This is one of the difficult steps of the proof. However, this boundedness follows directly from Theorem 4.2. The rest of the argument which modifies Δ to get a klt complement is an easy application of complement theory.

6 Some related problems and topics

6.1 Fano fibrations. One of the possible outcomes of the MMP is a Mori fibre space which is an extremal contraction $X \to Z$ where K_X is anti-ample over Z. This is a special kind of Fano fibration. Fano fibrations and more generally Fano type fibrations appear naturally in the course of applying induction on uniruled varieties, and in the context of moduli theory.

Suppose now that $f: X \to Z$ is a Mori fibre space where X is a 3-fold with $\mathbb{Q}$-factorial terminal singularities. Mori and Prokhorov proved that if Z is a surface, then Z has canonical singularities by Mori and Y. Prokhorov [2008], and if Z is a curve, then the coefficients of the fibres of f are bounded from above by 6 by Mori and Y. G. Prokhorov [2009].

M^c^Kernan proposed a generalisation of the first part to higher dimension:

Conjecture 6.2. *Assume $d \in \mathbb{N}$ and $\epsilon \in \mathbb{R}^{>0}$. Then there is $\delta \in \mathbb{R}^{>0}$ such that if $f: X \to Z$ is a Mori fibre space where X is ϵ-lc $\mathbb{Q}$-factorial of dimension d, then Z is δ-lc.*

On the other hand, independently, Shokurov proposed a more general problem which generalised both parts of Mori and Prokhorov result.

Conjecture 6.3. *Assume $d \in \mathbb{N}$ and $\epsilon \in \mathbb{R}^{>0}$. Then there is $\delta \in \mathbb{R}^{>0}$ such that if*

- *(X, B) is an ϵ-lc pair of dimension d,*
- *$f: X \to Z$ is a contraction with* $\dim Z > 0$*,*

- $K_X + B \sim_{\mathbb{R}} 0/Z$, *and* $-K_X$ *is big*$/Z$,

then we can write

$$K_X + B \sim_{\mathbb{R}} f^*(K_Z + B_Z + M_Z)$$

such that $(Z, B_Z + M_Z)$ *is* δ*-lc where* B_Z *and* M_Z *are the discriminant and moduli parts of adjunction.*

M^c^Kernan's conjecture is known in the toric case by V. Alexeev and A. Borisov [2014]. Shokurov's conjecture is known when $\dim X - \dim Z \leq 1$ by Birkar [2016c], in particular for surfaces, and open in higher dimension but we have the following general result of Birkar [ibid.].

Theorem 6.4. *Shokurov conjecture holds for those* f *such that* $(F, \operatorname{Supp} B|_F)$ *belongs to a bounded family where* F *is a general fibre of* f.

Note that by BAB (more precisely Birkar [2016b, Corollary 1.2]), F automatically belongs to a bounded family. However, one has little control over $\operatorname{Supp} B|_F$ and this is the main difficulty. This issue is similar to the difficulties which appear in the proof of BAB and related results. It is expected that the methods developed to prove BAB also works to prove Shokurov's conjecture but perhaps after some hard work.

6.5 Minimal log discrepancies and termination. The lc threshold plays an important role in birational geometry. This is clear from the proofs described in Section 5. It is also related to the termination conjecture (1.5) by Birkar [2007]. Another more subtle invariant of singularities is the *minimal log discrepancy* (mld) also defined by Shorkuov. Let (X, B) be a pair. The mld of (X, B) denoted $\operatorname{mld}(X, B)$ is defined to be the minimum of log discrepancies $a(D, X, B)$ where D runs over all prime divisors on birational model of X. The mld is way harder to treat than the lc threshold. Shokurov proposed the following:

Conjecture 6.6 (ACC for mld's). *Assume* $d \in \mathbb{N}$ *and* $\Phi \subset [0, 1]$ *is a set of numbers satisfying the descending chain condition (DCC). Then the set*

$$\{\operatorname{mld}(X, B) \mid (X, B) \text{ is an lc pair and coefficients of } B \text{ are in } \Phi\}$$

satisfies the ascending chain condition (ACC).

This is known for surfaces by V. Alexeev [1993] but open in dimension ≥ 3. Its importance is in relation with the termination conjecture and other topics of interest, see Shokurov [2004] and Birkar and Shokurov [2010]. Shokurov showed that this ACC conjecture together with a semi-continuity conjecture about mld's due to Ambro imply the termination conjecture by Shokurov [2004]. The expectation is that the ACC conjecture can be tackled using the theory of complements and the methods described in this text but again after some hard work.

6.7 Stable Fano varieties. Existence of specific metrics, e.g. Kähler-Einstein metrics, on manifolds is a central topic in differential geometry. Unlike canonically polarised and Calabi-Yau manifolds (see Yau [1978] and references therein), Fano manifolds do not always admit such metrics. It is now an established fact that a Fano manifold admits a Kähler-Einstein metric iff it is so-called *K-polystable* (see Chen, Donaldson, and Sun [2015] and references therein).

On the other hand, it is well-known that Fano varieties do not behave as well as canonically polarised varieties in the context of moduli theory. For example, the moduli space would not be separated. A remedy is to consider only *stable* Fano's. The first step of constructing a moduli space is to prove a suitable boundedness result. In the smooth case this is not an issue by Kollár, Miyaoka, and Mori [1992] but in the singular case boundedness is a recent result. Using methods described in Section 5, Jiang [2017] proved such a result by showing that the set of K-semistable Fano varieties X of fixed dimension and $\mathrm{vol}(-K_X)$ bounded from below forms a bounded family.

6.8 Other topics. There are connections between the advances described in this text and other topics of interest not discussed above. Here we only mention some works very briefly. Lehmann, Tanimoto, and Tschinkel [2014] and Lehmann and Tanimoto [2017] relate boundedness of Fano's and related invariants to the geometry underlying Manin's conjecture on distribution of rational points on Fano varieties. One the other hand, Cerbo and Svaldi [2016] studied boundedness of Calabi-Yau pairs where boundedness of Fano varieties appears naturally.

References

Valery Alexeev (1993). "Two two-dimensional terminations". *Duke Math. J.* 69.3, pp. 527–545. MR: 1208810 (cit. on p. 605).

– (1994). "Boundedness and K^2 for log surfaces". *Internat. J. Math.* 5.6, pp. 779–810. MR: 1298994 (cit. on p. 594).

Valery A. Alexeev and Michel Brion (2004). "Boundedness of spherical Fano varieties". In: *The Fano Conference*. Univ. Torino, Turin, pp. 69–80. MR: 2112568 (cit. on p. 594).

Valery Alexeev and Alexander Borisov (2014). "On the log discrepancies in toric Mori contractions". *Proc. Amer. Math. Soc.* 142.11, pp. 3687–3694. MR: 3251710 (cit. on p. 605).

Valery Alexeev and Shigefumi Mori (2004). "Bounding singular surfaces of general type". In: *Algebra, arithmetic and geometry with applications (West Lafayette, IN, 2000)*. Springer, Berlin, pp. 143–174. MR: 2037085 (cit. on p. 595).

Florin Ambro (2016). "Variation of log canonical thresholds in linear systems". *Int. Math. Res. Not. IMRN* 14, pp. 4418–4448. MR: 3556423 (cit. on p. 597).

C. Birkar and V. V. Shokurov (2010). "Mld's vs thresholds and flips". *J. Reine Angew. Math.* 638, pp. 209–234. MR: 2595341 (cit. on p. 605).

Caucher Birkar (2007). "Ascending chain condition for log canonical thresholds and termination of log flips". *Duke Math. J.* 136.1, pp. 173–180. MR: 2271298 (cit. on p. 605).

– (2010). "On existence of log minimal models". *Compos. Math.* 146.4, pp. 919–928. MR: 2660678 (cit. on p. 585).

– (2011). "On existence of log minimal models II". *J. Reine Angew. Math.* 658, pp. 99–113. MR: 2831514 (cit. on p. 585).

– (2012). "Existence of log canonical flips and a special LMMP". *Publ. Math. Inst. Hautes Études Sci.* 115, pp. 325–368. MR: 2929730 (cit. on p. 585).

– (Mar. 2016a). "Anti-pluricanonical systems on Fano varieties". arXiv: 1603.05765 (cit. on pp. 587, 590, 592, 594, 600, 602).

– (Sept. 2016b). "Singularities of linear systems and boundedness of Fano varieties". arXiv: 1609.05543 (cit. on pp. 587, 590, 592, 594, 596–598, 605).

– (2016c). "Singularities on the base of a Fano type fibration". *J. Reine Angew. Math.* 715, pp. 125–142. MR: 3507921 (cit. on p. 605).

Caucher Birkar, Paolo Cascini, Christopher D. Hacon, and James McKernan (2010). "Existence of minimal models for varieties of log general type". *J. Amer. Math. Soc.* 23.2, pp. 405–468. MR: 2601039 (cit. on pp. 585, 587, 588).

Caucher Birkar and De-Qi Zhang (2016). "Effectivity of Iitaka fibrations and pluricanonical systems of polarized pairs". *Publ. Math. Inst. Hautes Études Sci.* 123, pp. 283–331. MR: 3502099 (cit. on pp. 587, 590, 599).

A. A. Borisov and L. A. Borisov (1992). "Singular toric Fano three-folds". *Mat. Sb.* 183.2, pp. 134–141. MR: 1166957 (cit. on pp. 594, 603).

Alexandr Borisov (2001). "Boundedness of Fano threefolds with log-terminal singularities of given index". *J. Math. Sci. Univ. Tokyo* 8.2, pp. 329–342. MR: 1837167 (cit. on p. 594).

Gabriele Di Cerbo and Roberto Svaldi (Aug. 2016). "Birational boundedness of low dimensional elliptic Calabi-Yau varieties with a section". arXiv: 1608.02997 (cit. on p. 606).

I. A. Cheltsov and K. A. Shramov (2008). "Log-canonical thresholds for nonsingular Fano threefolds". *Uspekhi Mat. Nauk* 63.5(383). Appendix by J.-P. Demailly: *On Tian's invariant and log canonical thresholds*, pp. 73–180. MR: 2484031 (cit. on p. 597).

Xiuxiong Chen, Simon Donaldson, and Song Sun (2015). "Kähler-Einstein metrics on Fano manifolds. III: Limits as cone angle approaches 2π and completion of the main proof". *J. Amer. Math. Soc.* 28.1, pp. 235–278. MR: 3264768 (cit. on p. 606).

Christopher D. Hacon and James McKernan (2006). "Boundedness of pluricanonical maps of varieties of general type". *Invent. Math.* 166.1, pp. 1–25. MR: 2242631 (cit. on p. 586).

– (2010). "Existence of minimal models for varieties of log general type. II". *J. Amer. Math. Soc.* 23.2, pp. 469–490. MR: 2601040 (cit. on p. 585).

Christopher D. Hacon, James McKernan, and Chenyang Xu (2013). "On the birational automorphisms of varieties of general type". *Ann. of Math. (2)* 177.3, pp. 1077–1111. MR: 3034294 (cit. on p. 586).

– (2014). "ACC for log canonical thresholds". *Ann. of Math. (2)* 180.2, pp. 523–571. MR: 3224718 (cit. on pp. 586, 594, 599).

Christopher D. Hacon and Chenyang Xu (2015). "Boundedness of log Calabi-Yau pairs of Fano type". *Math. Res. Lett.* 22.6, pp. 1699–1716. arXiv: 1410.8187. MR: 3507257 (cit. on pp. 594, 603).

Chen Jiang (Nov. 2014). "Boundedness of anti-canonical volumes of singular log Fano threefolds". arXiv: 1411.6728 (cit. on p. 597).

– (Sept. 2015). "On birational boundedness of Fano fibrations". arXiv: 1509.08722 (cit. on pp. 594, 597).

– (May 2017). "Boundedness of $\mathbb{Q}$-Fano varieties with degrees and alpha-invariants bounded from below". arXiv: 1705.02740 (cit. on p. 606).

Yujiro Kawamata (1984). "Elementary contractions of algebraic 3-folds". *Ann. of Math. (2)* 119.1, pp. 95–110. MR: 736561 (cit. on p. 585).

– (1992a). "Abundance theorem for minimal threefolds". *Invent. Math.* 108.2, pp. 229–246. MR: 1161091 (cit. on p. 585).

– (1992b). "Boundedness of **Q**-Fano threefolds". In: *Proceedings of the International Conference on Algebra, Part 3 (Novosibirsk, 1989)*. Vol. 131. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 439–445. MR: 1175897 (cit. on p. 594).

– (1992c). "Termination of log flips for algebraic 3-folds". *Internat. J. Math.* 3.5, pp. 653–659. MR: 1189678 (cit. on p. 585).

– (1998). "Subadjunction of log canonical divisors. II". *Amer. J. Math.* 120.5, pp. 893–899. MR: 1646046 (cit. on p. 600).

János Kollár, Yoichi Miyaoka, and Shigefumi Mori (1992). "Rational connectedness and boundedness of Fano manifolds". *J. Differential Geom.* 36.3, pp. 765–779. MR: 1189503 (cit. on pp. 594, 606).

János Kollár, Yoichi Miyaoka, Shigefumi Mori, and Hiromichi Takagi (2000). "Boundedness of canonical **Q**-Fano 3-folds". *Proc. Japan Acad. Ser. A Math. Sci.* 76.5, pp. 73–77. MR: 1771144 (cit. on p. 594).

János Kollár and Shigefumi Mori (1992). "Classification of three-dimensional flips". *J. Amer. Math. Soc.* 5.3, pp. 533–703. MR: 1149195 (cit. on p. 593).

– (1998). *Birational geometry of algebraic varieties*. Vol. 134. Cambridge Tracts in Mathematics. With the collaboration of C. H. Clemens and A. Corti, Translated from the 1998 Japanese original. Cambridge University Press, Cambridge, pp. viii+254. MR: 1658959 (cit. on pp. 588, 590).

Brian Lehmann and Sho Tanimoto (Feb. 2017). "Geometric Manin's Conjecture and rational curves". arXiv: 1702.08508 (cit. on p. 606).

Brian Lehmann, Sho Tanimoto, and Yuri Tschinkel (Sept. 2014). "Balanced line bundles on Fano varieties". To appear in *J. Reine Angew. Math.* arXiv: 1409.5901 (cit. on p. 606).

Yoichi Miyaoka (1988). "Abundance conjecture for 3-folds: case $\nu = 1$". *Compositio Math.* 68.2, pp. 203–220. MR: 966580 (cit. on p. 585).

S. Mori and Yu. G. Prokhorov (2009). "Multiple fibers of del Pezzo fibrations". *Tr. Mat. Inst. Steklova* 264.Mnogomernaya Algebraicheskaya Geometriya, pp. 137–151. MR: 2590844 (cit. on p. 604).

Shigefumi Mori (1988). "Flip theorem and the existence of minimal models for 3-folds". *J. Amer. Math. Soc.* 1.1, pp. 117–253. MR: 924704 (cit. on pp. 585, 593).

Shigefumi Mori and Yuri Prokhorov (2008). "On $\mathbb{Q}$-conic bundles". *Publ. Res. Inst. Math. Sci.* 44.2, pp. 315–369. MR: 2426350 (cit. on p. 604).

Yu. G. Prokhorov and V. V. Shokurov (2001). "The first fundamental theorem on complements: from global to local". *Izv. Ross. Akad. Nauk Ser. Mat.* 65.6, pp. 99–128. MR: 1892905 (cit. on p. 592).

– (2009). "Towards the second main theorem on complements". *J. Algebraic Geom.* 18.1, pp. 151–199. MR: 2448282 (cit. on p. 592).

Yuri Prokhorov and Constantin Shramov (2016). "Jordan property for Cremona groups". *Amer. J. Math.* 138.2, pp. 403–418. MR: 3483470 (cit. on p. 596).

M. Reid (n.d.). "Projective morphisms according to Kawamata". Preprint from Univ. of Warwick, Coventry, (1983) (cit. on p. 593).

Miles Reid (1987). "Young person's guide to canonical singularities". In: *Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985)*. Vol. 46. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 345–414. MR: 927963 (cit. on p. 593).

Jean-Pierre Serre (2009). "A Minkowski-style bound for the orders of the finite subgroups of the Cremona group of rank 2 over an arbitrary field". *Mosc. Math. J.* 9.1, 193–208, back matter. MR: 2567402 (cit. on p. 596).

V. V. Shokurov (1985). "A nonvanishing theorem". *Izv. Akad. Nauk SSSR Ser. Mat.* 49.3, pp. 635–651. MR: 794958 (cit. on p. 585).

– (1993). "3-fold log flips". *Izv. Ross. Akad. Nauk Ser. Mat.* 40.1. With an appendix by Yujiro Kawamata, pp. 95–202 (cit. on pp. 585, 592).

– (2000). "Complements on surfaces". *J. Math. Sci. (New York)* 102.2. Algebraic geometry, 10, pp. 3876–3932. MR: 1794169 (cit. on pp. 592, 593).

V. V. Shokurov (2003). “Prelimiting flips”. *Tr. Mat. Inst. Steklova* 240.Biratsion. Geom. Lineĭn. Sist. Konechno Porozhdennye Algebry, pp. 82–219. MR: 1993750 (cit. on p. 585).

– (2004). “Letters of a bi-rationalist. V. Minimal log discrepancies and termination of log flips”. *Tr. Mat. Inst. Steklova* 246.Algebr. Geom. Metody, Svyazi i Prilozh. Pp. 328–351. MR: 2101303 (cit. on p. 605).

– (2009). “Letters of a bi-rationalist. VII. Ordered termination”. *Tr. Mat. Inst. Steklova* 264.Mnogomernaya Algebraicheskaya Geometriya, pp. 184–208. MR: 2590847 (cit. on p. 585).

V. V. Šokurov (1979). “Smoothness of a general anticanonical divisor on a Fano variety”. *Izv. Akad. Nauk SSSR Ser. Mat.* 43.2, pp. 430–441. MR: 534602 (cit. on p. 593).

Shigeharu Takayama (2006). “Pluricanonical systems on algebraic varieties of general type”. *Invent. Math.* 165.3, pp. 551–587. MR: 2242627 (cit. on p. 586).

Gang Tian (1990). “On a set of polarized Kähler metrics on algebraic manifolds”. *J. Differential Geom.* 32.1, pp. 99–130. MR: 1064867 (cit. on p. 597).

Shing Tung Yau (1978). “On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I”. *Comm. Pure Appl. Math.* 31.3, pp. 339–411. MR: 480350 (cit. on p. 606).

Received 2017-12-05.

Caucher Birkar
c.birkar@dpmms.cam.ac.uk

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (609–636)

VARIATIONAL AND NON-ARCHIMEDEAN ASPECTS OF THE YAU–TIAN–DONALDSON CONJECTURE

Sébastien Boucksom

Abstract

We survey some recent developments in the direction of the Yau-Tian-Donaldson conjecture, which relates the existence of constant scalar curvature Kähler metrics to the algebro-geometric notion of K-stability. The emphasis is put on the use of pluripotential theory and the interpretation of K-stability in terms of non-Archimedean geometry.

Contents

Introduction

The search for constant curvature metrics is a recurring theme in geometry, the fundamental uniformization theorem for Riemann surfaces being for instance equivalent to the existence of a (complete) Hermitian metric with constant curvature on any one-dimensional complex manifold. On a higher dimensional complex manifold, *Kähler metrics* are defined as Hermitian metrics locally expressed as the complex Hessian of some (plurisubharmonic) function, known as a *local potential* for the metric. As a result, constant curvature problems for Kähler metrics boil down to scalar PDEs for their potentials, a famous

The author was partially supported by the ANR project GRACK..
MSC2010: primary 58E11; secondary 26E30, 14L24, 32U5.

instance being Kähler metrics with constant Ricci curvature, known as *Kähler-Einstein* metrics, whose local potentials satisfy a complex Monge-Ampère equation. This was in fact a main motivation for the introduction of Kähler metrics in Kähler [1933], where it was also noted that the complex Monge-Ampère equation in question can be written as the Euler-Lagrange equation of a certain functional.

In the present paper, we will more generally consider *constant scalar curvature* Kähler metrics (cscK metrics for short) on a compact complex manifold X. Kähler metrics in a fixed cohomology class of X are parametrized by a space $\mathcal{H}$ of (global) Kähler potentials $u \in C^\infty(X)$, cscK metrics corresponding to solutions in $\mathcal{H}$ of a certain fourth-order nonlinear elliptic PDE. Remarkably, the latter is again the Euler-Lagrange equation of a functional M on $\mathcal{H}$, discovered by T. Mabuchi. While M is generally not convex on $\mathcal{H}$ as an open convex subset of $C^\infty(X)$, Mabuchi defined a natural Riemannian L^2-metric on $\mathcal{H}$ with respect to which M does become convex, opening the way to a variational approach to the cscK problem. The picture was further clarified by S.K. Donaldson, who noted that $\mathcal{H}$ behaves like an infinite dimensional symmetric space and emphasized the analogy with the log norm function in Geometric Invariant Theory.

Using this as a guide, one would like to detect the growth properties of M by looking at its slope at infinity along certain geodesic rays in $\mathcal{H}$ arising from algebro-geometric one-parameter subgroups, and prove that positivity of these slopes ensures the existence of a minimizer, which would then be a cscK metric. This is basically the prediction of the *Yau-Tian-Donaldson conjecture*, positivity of the algebro-geometric slopes at infinity being equivalent to *K-stability*. In the Kähler-Einstein case, this conjecture was famously solved a few years ago by Chen, S. Donaldson, and Sun [2015a,b,c], thereby completing intensive research on positively curved Kähler-Einstein metrics with many key contributions by G.Tian.

The more elementary case of convex functions on (finite dimensional) Riemannian symmetric spaces (see Section 1.3) and experience from the direct method of the calculus of variations suggest to try to attack the general case of the conjecture along the following steps:

1. extend M to a convex functional on a certain metric completion $\bar{\mathcal{H}}$, in which coercivity (i.e. linear growth) implies the existence of a minimizer;

2. prove that a minimizer u of M in $\bar{\mathcal{H}}$ is a weak solution to the cscK PDE in some appropriate sense, and show that ellipticity of this equation implies that u is smooth, hence a cscK potential;

3. show that M is either coercive, or bounded above on some geodesic ray in $\bar{\mathcal{H}}$;

4. approximate any geodesic ray (u_t) in $\bar{\mathcal{H}}$ by algebro-geometric rays $(u_{j,t})$ in $\mathcal{H}$, in such a way that (uniform) positivity of the slopes of M along $(u_{j,t})$ forces $M(u_t) \to +\infty$ at infinity.

As of this writing, (1) and (3) are fully understood, as a combination of Chen [2000b], Darvas [2015], Darvas and Rubinstein [2017], Berman and Berndtsson [2017], Berman, Boucksom, Eyssidieux, Guedj, and Zeriahi [2011], and Berman, Boucksom, and Jonsson [2015]. On the other hand, while (2) and (4) are known in the Kähler-Einstein case Berman, Boucksom, Eyssidieux, Guedj, and Zeriahi [2011] and Berman, Boucksom, and Jonsson [2015], they remain wide open in general[1]. The goal of this text is to survey these developments, as well as the analysis of the algebro-geometric slopes at infinity in terms of non-Archimedean geometry, building on Kontsevich and Tschinkel [2000] and Boucksom, Favre, and Jonsson [2016, 2015]. It is organized as follows:

- Section 1 describes the 'baby case' of convex functions on the space of Hermitian norms of a fixed vector space, introducing alternative Finsler metrics and the space of non-Archimedean norms as the cone at infinity;
- Section 2 recalls the basic formalism of Kähler potentials and energy functionals;
- Section 3 reviews the link between the metric geometry of $\mathcal{H}$ and pluripotential theory, and discusses (1), (2) and (3) above;
- Section 4 introduces the non-Archimedean counterparts to Kähler potentials and the energy functionals, and presents a proof of (4) in the Kähler-Einstein case.

Acknowledgments. My current view of the subject has been framed by collaborations and countless discussions with many mathematicians. Without attempting to be exhaustive, I would like to thank in particular Bo Berndtsson, Jean-Pierre Demailly, Ruadhaí Dervan, Dennis Eriksson, Philippe Eyssidieux, Paul Gauduchon, Henri Guenancia, Tomoyuki Hisamoto, Mihai Păun, Valentino Tosatti and David Witt Nyström, and to express my profound gratitude to Robert Berman, Charles Favre, Vincent Guedj, Mattias Jonsson and Ahmed Zeriahi for the key role their ideas have been playing in our joint works over the years.

1 Convex functions on spaces of norms

The complexification G of any compact Lie group K is a reductive complex algebraic group, giving rise to a Riemannian symmetric space G/K and a conical Tits building.

[1] A proof of (2) has recently been announced by Chen and Cheng [2017, 2018a,b]

The latter can be viewed as the asymptotic cone of G/K, and the growth properties of any convex, Lipschitz continuous function on G/K are encoded in an induced function on the building. While this picture is well-known (see for instance Kapovich, Leeb, and Millson [2009]), it becomes very explicit for the unitary group $U(N)$, for which $G/K \simeq \mathcal{N}$ is the space of Hermitian norms on $\mathbb{C}^N$. The goal of this section is to discuss this case in elementary terms, along with alternative Finsler metrics on $\mathcal{N}$, providing a finite dimensional version of the more sophisticated Kähler geometric setting considered afterwards.

1.1 Finsler geometry on the space of norms. Let V be a complex vector space of finite dimension N, and denote by $\mathcal{N}$ the space of Hermitian norms γ on V, viewed as an open subset of the (N^2-dimensional) real vector space $\mathrm{Herm}(V)$ of Hermitian forms h. The ordered spectrum of $h \in \mathrm{Herm}(V)$ with respect to $\gamma \in \mathcal{N}$ defines a point $\lambda_\gamma(h)$ in the *Weyl chamber*

$$\mathcal{C} = \{\lambda \in \mathbb{R}^N \mid \lambda_1 \geq \cdots \geq \lambda_N\} \simeq \mathbb{R}^N/\mathfrak{S}_N,$$

where the symmetric group $\mathfrak{S}_N$ acts on $\mathbb{R}^N$ by permuting coordinates.

Lemma 1.1. *For each symmetric* (i.e. $\mathfrak{S}_N$*-invariant*) *norm* χ *on* $\mathbb{R}^N$*, we have*

$$\chi\left(\lambda_\gamma(h+h')\right) \leq \chi\left(\lambda_\gamma(h)\right) + \chi\left(\lambda_\gamma(h')\right)$$

for all $\gamma \in \mathcal{N}$ *and* $h, h' \in \mathrm{Herm}(V)$.

Proof. Given $\lambda, \lambda' \in \mathcal{C}$, one says that λ is *majorized by* λ', written $\lambda \preceq \lambda'$, if

$$\lambda_1 + \cdots + \lambda_i \leq \lambda'_1 + \cdots + \lambda'_i$$

for all i, with equality for $i = N$. It is a well-known and simple consequence of the Hahn–Banach theorem that $\lambda \preceq \lambda'$ iff λ is in the convex envelope of the $\mathfrak{S}_N$-orbit of λ', which implies $\chi(\lambda) \leq \chi(\lambda')$ by convexity, homogeneity and $\mathfrak{S}_N$-invariance of χ. The Lemma now follows from the classical Ky Fan inequality $\lambda_\gamma(h+h') \preceq \lambda_\gamma(h) + \lambda_\gamma(h')$. □

Thanks to Lemma 1.1, setting $|h|_{\chi,\gamma} := \chi(\lambda_\gamma(h))$ defines a continuous Finsler norm $|\cdot|_\chi$ on $\mathcal{N}$, and hence a length metric d_χ on $\mathcal{N}$, with $d_\chi(\gamma,\gamma')$ defined as usual as the infimum of the lengths $\int_0^1 |\dot{\gamma}_t|_{\chi,\gamma_t}\,dt$ of all smooth paths $(\gamma_t)_{t\in[0,1]}$ in $\mathcal{N}$ joining γ to γ'. By equivalence of norms in $\mathbb{R}^N$, all metrics d_χ on $\mathcal{N}$ are Lipschitz equivalent.

Example 1.2. *The metric* d_2 *induced by the* ℓ^2*-norm on* $\mathbb{R}^N$ *is the usual Riemannian metric of* $\mathcal{N}$ *identified with the Riemannian symmetric space* $\mathrm{GL}(N,\mathbb{C})/U(N)$*. In particular,* $(\mathcal{N}, d_2)$ *is a complete* $\mathrm{CAT}(0)$*-space, a nonpositive curvature condition implying that any two points of* $\mathcal{N}$ *are joined by a unique (length minimizing) geodesic.*

Example 1.3. *The metric d_∞ induced by the ℓ^∞-norm on $\mathbb{R}^N$ admits a direct description as a sup-norm*

$$d_\infty(\gamma,\gamma') = \sup_{v\in V\setminus\{0\}} \left|\log\gamma(v) - \log\gamma'(v)\right|,$$

whose exponential is the best constant $C > 0$ such that $C^{-1}\gamma \le \gamma' \le C\gamma$ on V.

In order to describe the geometry of $(\mathfrak{N}, d_\chi)$, introduce for each basis $e = (e_1,\dots,e_N)$ of V the embedding

$$\iota_e : \mathbb{R}^N \hookrightarrow \mathfrak{N}$$

that sends $\lambda \in \mathbb{R}^N$ to the Hermitian norm for which e is orthogonal and e_i has norm $e^{-\lambda_i}$. The image $\iota_e(\mathbb{R}^N)$ is thus the set of norms in $\mathfrak{N}$ that are diagonalized in the given basis e. Any two $\gamma,\gamma' \in \mathfrak{N}$ can be jointly diagonalized in some basis e, i.e. $\gamma = \iota_e(\lambda)$, $\gamma' = \iota_e(\lambda')$ with $\lambda,\lambda' \in \mathbb{R}^N$. After permutation, the vector $\lambda'-\lambda$ determines an element $\lambda(\gamma,\gamma') \in \mathfrak{C}$ which only depends on γ,γ', and is obtained by applying $-\log$ to the spectrum of γ' with respect to γ. The following result, proved in Boucksom and Eriksson [2018], generalizes the well-known Riemannian picture for d_2.

Theorem 1.4. *For each symmetric norm χ on $\mathbb{R}^N$, the induced Finsler metric d_χ on $\mathfrak{N}$ is given by $d_\chi(\gamma,\gamma') = \chi(\lambda(\gamma,\gamma'))$ for all $\gamma,\gamma' \in \mathfrak{N}$. It is further characterized as the unique metric on $\mathfrak{N}$ such that $\iota_e : (\mathbb{R}^N,\chi) \hookrightarrow (\mathfrak{N}, d_\chi)$ is an isometric embedding for all bases e.*

1.2 Convergence to non-Archimedean norms. By a *geodesic ray* $(\gamma_t)_{t\in\mathbb{R}_+}$ in $\mathfrak{N}$, we mean a constant speed Riemannian geodesic ray, i.e. $d_2(\gamma_t,\gamma_s)$ is a constant multiple of $|t-s|$. Every geodesic ray is of the form $\gamma_t = \iota_e(t\lambda)$ for some basis e and $\lambda \in \mathbb{R}^N$, the latter being uniquely determined up to permutation as the spectrum of the Hermitian form $\dot\gamma_t$ with respect to γ_t for any value of t. As a result, (γ_t) is also a (constant speed) geodesic ray for all Finsler metrics d_χ, and indeed satisfies $d_\chi(\gamma_t,\gamma_s) = \chi(\lambda)|t-s|$. The metric d_χ might admit other geodesic rays in general, but we will not consider these in what follows.

Two geodesic rays (γ_t), (γ'_t) are called *asymptotic* if γ_t and γ'_t stay at bounded distance with respect to any of the Lipschitz equivalent metrics d_χ, i.e. are uniformly equivalent as norms on V. This defines an equivalence relation on the set of geodesic rays, whose quotient naturally identifies with a space of *non-Archimedean norms*.

To see this, pick a geodesic ray $\gamma_t = \iota_e(t\lambda)$. Then $\gamma_t(v)^2 = \sum_i |v_i|^2 e^{-2\lambda_i t}$ for each vector $v = \sum_i v_i e_i$ in V, from which one easily gets that $\gamma_t(v)^{1/t}$ converges to

$$\alpha\left(\sum_i v_i e_i\right) := \max_{v_i\neq 0} e^{-\lambda_i}. \tag{1-1}$$

as $t \to \infty$. The function $\alpha : V \to \mathbb{R}_+$ so defined satisfies

(i) $\alpha(v + v') \le \max\{\alpha(v), \alpha(v')\}$;

(ii) $\alpha(\tau v) = \alpha(v)$ for all $\tau \in \mathbb{C}^*$;

(iii) $\alpha(v) = 0 \iff v = 0$,

which means that α is an element of the space $\mathfrak{N}^{\mathrm{NA}}$ of non-Archimedean norms on V with respect to the *trivial absolute value* $|\cdot|_0$ on the ground field $\mathbb{C}$, i.e. $|0|_0 = 0$ and $|\tau|_0 = 1$ for $\tau \in \mathbb{C}^*$. The closed balls of such a norm are linear subspaces of V, and the data of α thus amounts to that of an $\mathbb{R}$-filtration of V, or equivalently a flag of linear subspaces together with a tuple of real numbers; for this reason, $\mathfrak{N}^{\mathrm{NA}}$ is also known in the literature as the *(conical) flag complex*. The space $\mathfrak{N}^{\mathrm{NA}}$ has a natural $\mathbb{R}_+^*$-action $(t, \alpha) \mapsto \alpha^t$, whose only fixed point is the *trivial norm* α_0 on V.

The existence of a basis of V compatible with a given flag implies that any non-Archimedean norm $\alpha \in \mathfrak{N}^{\mathrm{NA}}$ can be diagonalized in some basis $e = (e_i)$, in the sense that it satisfies Equation (1-1) for some $\lambda \in \mathbb{R}^N$. The image of λ in $\mathbb{R}^N/\mathfrak{S}_N$ is uniquely determined by α, and a complete invariant for the (non-transitive) action of $G = \mathrm{GL}(V)$ on $\mathfrak{N}^{\mathrm{NA}}$, inducing an identification

$$\mathfrak{N}^{\mathrm{NA}}/G \simeq \mathbb{R}^N/\mathfrak{S}_N.$$

The structure of $\mathfrak{N}^{\mathrm{NA}}$ can be analyzed just as that of $\mathfrak{N}$ by introducing for each basis e the embedding

$$\iota_e^{\mathrm{NA}} : \mathbb{R}^N \hookrightarrow \mathfrak{N}^{\mathrm{NA}}$$

sending $\lambda \in \mathbb{R}^N$ to the non-Archimedean norm (Equation (1-1)). Any two norms can be jointly diagonalized, i.e. belong to the image of ι_e for some e, and it is proved in Boucksom and Eriksson [2018] that there exists a unique metric d_χ^{NA} on $\mathfrak{N}^{\mathrm{NA}}$ for which each $\iota_e^{\mathrm{NA}} : (\mathbb{R}^N, \chi) \to (\mathfrak{N}^{\mathrm{NA}}, d_\chi^{\mathrm{NA}})$ is an isometric embedding. It is worth mentioning that the Lipschitz equivalent metric spaces $(\mathfrak{N}^{\mathrm{NA}}, d_\chi^{\mathrm{NA}})$, while complete, are *not* locally compact as soon as $N > 1$.

Example 1.5. *Every (algebraic)* 1*-parameter subgroup* $\rho : \mathbb{C}^* \to \mathrm{GL}(V)$ *defines a non-Archimedean norm* $\alpha_\rho \in \mathfrak{N}^{\mathrm{NA}}$*, characterized by*

$$\alpha_\rho(v) \le r \iff \lim_{\tau \to 0} \tau^{\lceil \log r \rceil} \rho(\tau) \cdot v \text{ exists in } V.$$

If $e = (e_i)$ *is a basis of eigenvectors for* ρ *with* $\rho(\tau) \cdot e_i = \tau^{\lambda_i} e_i$, $\lambda_i \in \mathbb{Z}$, *then* $\alpha_\rho = \iota_e(\lambda)$. *This shows that the lattice points* $\mathfrak{N}_{\mathbb{Z}}^{\mathrm{NA}}$, i.e. *the images of* $\mathbb{Z}^N$ *by the embeddings* ι_e*, are exactly the norms attached to* 1*-parameter subgroups, and ultimately leads to an identification of* $(\mathfrak{N}^{\mathrm{NA}}, d_2)$ *with the (conical) Tits building of the reductive algebraic group* $\mathrm{GL}(V)$.

Coming back to geodesic rays, one proves that the non-Archimedean norms $\alpha = \lim \gamma_t^{1/t}$, $\alpha' = \lim \gamma_t'^{1/t}$ defined by two rays (γ_t), (γ_t') are equal iff the rays are asymptotic, and that d_χ^{NA} computes the slope at infinity of d_χ, i.e.

$$d_\chi^{\mathrm{NA}}(\alpha, \alpha') = \lim_{t\to\infty} \frac{d_\chi(\gamma_t, \gamma_t')}{t}. \tag{1-2}$$

1.3 Slopes at infinity of a convex function. If $f : \mathbb{R}_+ \to \mathbb{R}$ is convex, $(f(t) - f(0))/t$ is a nondecreasing function of t. The *slope at infinity*

$$f'(\infty) := \lim_{t\to+\infty} \frac{f(t)}{t} \in (-\infty, +\infty]$$

is thus well-defined, and finite if f is Lipschitz continuous. It is characterized as the supremum of all $s \in \mathbb{R}$ such that $f(t) \geq st + O(1)$ on $\mathbb{R}_+$, and f is bounded above iff $f'(\infty) \leq 0$.

A function $F : \mathfrak{N} \to \mathbb{R}$ on the space of Hermitian norms is (geodesically) convex iff $F \circ \iota_e : \mathbb{R}^N \to \mathbb{R}$ is convex for each basis e, and similarly for a function on $\mathfrak{N}^{\mathrm{NA}}$. Assume further that F is Lipschitz. Then $F(\gamma_t)$ is convex and Lipschitz continuous on $\mathbb{R}_+$ for each geodesic ray γ, and the slope at infinity $\lim_{t\to+\infty} F(\gamma_t)/t$ only depends on the equivalence class $\alpha \in \mathfrak{N}^{\mathrm{NA}}$ defined by γ. As a result, F determines a function

$$F^{\mathrm{NA}} : \mathfrak{N}^{\mathrm{NA}} \to \mathbb{R},$$

characterized by $F(\gamma_t)/t \to F^{\mathrm{NA}}(\alpha)$ for each ray (γ_t) asymptotic to $\alpha \in \mathfrak{N}^{\mathrm{NA}}$, and this function is further convex and Lipschitz continuous by Equation (1-2).

Theorem 1.6. *Let $F : \mathfrak{N} \to \mathbb{R}$ be a convex, Lipschitz continuous function, and fix a base point $\gamma_0 \in \mathfrak{N}$ and a symmetric norm χ on $\mathbb{R}^N$. The following are equivalent:*

(i) $F : \mathfrak{N} \to \mathbb{R}$ is an exhaustion function, i.e. *proper and bounded below;*

(ii) F is coercive, i.e. $F(\gamma) \geq \delta\, d_\chi(\gamma, \gamma_0) - C$ *for some constants $\delta, C > 0$;*

(iii) $F^{\mathrm{NA}}(\alpha) > 0$ for all nontrivial $\alpha \in \mathfrak{N}^{\mathrm{NA}}$;

(iv) there exists $\delta > 0$ such that $F^{\mathrm{NA}} \geq \delta\, d_\chi^{\mathrm{NA}}$.

These conditions are further satisfied as soon as F admits a unique minimizer.

Proof. Clearly, (ii) implies (i), and (i) implies that $F(\gamma_t)$ is unbounded for any geodesic ray, hence has a positive slope at infinity, which yields (iii). Let us now prove (iii)$\Longrightarrow$(ii). Assuming by contradiction that there exists a sequence γ_j in $\mathfrak{N}$ such that

$$F(\gamma_j) \leq \delta_j d_\chi(\gamma_j, \gamma_0) - C_j \tag{1-3}$$

with $\delta_j \to 0$ and $C_j \to +\infty$, we are going to construct a non-constant geodesic ray (γ_t) along which F is bounded above, contradicting the positivity of the slope at infinity along this ray. By Lipschitz continuity, Equation (1-3) implies $T_j := d_\chi(\gamma_j, \gamma_0) \to \infty$. For each j, let $(\gamma_{j,t})_{t\in[0,T_j]}$ be the geodesic segment joining γ_0 to γ_j, parametrized so that $t = d_\chi(\gamma_{j,t}, \gamma_0)$. By Ascoli's theorem, $(\gamma_{j,t})$ converges to a geodesic ray (γ_t), uniformly on compact sets of $\mathbb{R}_+$. By convexity of F, we have

$$\frac{F(\gamma_{j,t}) - F(\gamma_0)}{t} \leq \frac{F(\gamma_j) - F(\gamma_0)}{T_j},$$

hence $F(\gamma_{j,t}) \leq \delta_j t + F(\gamma_0)$, which yields in the limit the upper bound $F(\gamma_t) \leq F(\gamma_0)$. At this point, we have thus shown that (i), (ii) and (iii) are equivalent. That (ii)$\Longrightarrow$(iv) follows from Equation (1-2), while (iv) clearly implies (iii).

Assume finally that F admits a unique minimizer, which we may take as the base point γ_0. If F is not coercive, the previous argument yields a nonconstant ray (γ_t) such that $F(\gamma_t) \leq F(\gamma_0) = \inf F$, which shows that all γ_t are mininizers of F, and hence $\gamma_t = \gamma_0$ by uniqueness, a contradiction. □

2 The constant scalar curvature problem for Kähler metrics

This section recalls the basic formalism of constant curvature Kähler metrics, and introduces the corresponding energy functionals.

2.1 Kähler metrics with constant curvature. Let X be a compact complex manifold, and denote by n its (complex) dimension. The data of a Hermitian metric on the tangent bundle T_X is equivalent to that of a positive $(1,1)$-form ω, locally expressed in holomorphic coordinates (z_j) as $\omega = \sqrt{-1}\sum_{ij} \omega_{ij}\, dz_i \wedge d\bar{z}_j$ with (ω_{ij}) a smooth family of positive definite Hermitian matrices. One says that ω is *Kähler* if it satisfies the following equivalent conditions:

(i) $d\omega = 0$;

(ii) ω admits local potentials, i.e. smooth real valued valued functions u such that $\omega = \sqrt{-1}\partial\bar{\partial}u$, or $\omega_{ij} = \partial^2 u/\partial z_i \partial \bar{z}_j$ in local coordinates;

(iii) the Levi-Civita connection ∇ of ω on the tangent bundle T_X coincides with the Chern connection, i.e. the unique Hermitian connection with $\nabla^{0,1} = \bar{\partial}$.

The Kähler condition thus ensures compatibility between Riemannian and complex Hermitian geometry. The (normalized) curvature tensor $\Theta_\omega(T_X) := \frac{\sqrt{-1}}{2\pi}\nabla^2$ of a Kähler metric is a $(1,1)$-form with values in the Hermitian endomorphisms of T_X, whose trace with respect to T_X coincides with the *Ricci curvature* $\mathrm{Ric}(\omega)$ in the sense of Riemannian geometry. In other words, the Ricci tensor of a Kähler metric can be seen as the curvature of the induced metric on the dual of the *canonical bundle* $K_X := \det T_X^\star$, the factor 2π being included in the curvature so that the de Rham cohomology class of the closed $(1,1)$-form $\mathrm{Ric}(\omega)$ coincides with the first Chern class

$$c_1(X) := c_1(T_X) = -c_1(K_X).$$

In terms of the normalized operator $\mathrm{dd^c} := \frac{\sqrt{-1}}{2\pi}\partial\bar\partial$ and a local function u with $\omega = \mathrm{dd^c}\, u$, we have

$$\mathrm{Ric}(\omega) = -\,\mathrm{dd^c} \log\det\left(\frac{\partial^2 u}{\partial z_j \partial \bar z_k}\right),$$

which accounts for the ubiquity of the *complex Monge-Ampère operator* $u \mapsto \det\left(\partial^2 u/\partial z_j \partial \bar z_k\right)$ in Kähler geometry. Taking the trace of $\mathrm{Ric}(\omega)$ with respect to ω yields the *scalar curvature*

$$S(\omega) = n\frac{\mathrm{Ric}(\omega)\wedge\omega^{n-1}}{\omega^n} = \Delta \log\det\left(\frac{\partial^2 u}{\partial z_j \partial \bar z_k}\right).$$

Denote by $V := \int_X \omega^n = [\omega]^n$ the volume of ω, and observe that the mean value of $S(\omega)$ is the cohomological constant

$$V^{-1}\int_X S(\omega)\omega^n = nV^{-1}\int_X \mathrm{Ric}(\omega)\wedge\omega^{n-1} = -n\lambda$$

with

$$\lambda := V^{-1}\left(c_1(K_X)\cdot[\omega]^{n-1}\right).$$

As a result, there exists a unique function $\rho \in C^\infty(X)$, the *Ricci potential* of ω, such that

$$\begin{cases} \Delta\rho = S(\omega) + n\lambda \\ \int_X e^\rho \omega^n = 1. \end{cases}$$

This defines a smooth, positive probability measure $\mu_0 := e^\rho\omega^n$ which we call the *Ricci normalized volume form of* ω.

To the above three notions of curvature correspond the following three versions of the constant curvature problem.

(a) Requiring the full curvature tensor of ω to be constant, i.e.

$$\Theta_\omega(T_X) = -\frac{\lambda}{n}\omega\otimes \mathrm{id}_{T_X},$$

is a very strong condition which implies uniformization, in the sense that (X, ω) must be isomorphic (after scaling the metric) to the complex projective space ($\lambda < 0$), a finite quotient of a compact complex torus ($\lambda = 0$), or a cocompact quotient of the complex hyperbolic ball ($\lambda > 0$).

(b) A *Kähler-Einstein metric* (KE for short) is a Kähler metric ω of constant Ricci curvature, i.e. satisfying $\mathrm{Ric}(\omega) = -\lambda\omega$, the Kähler analogue of the Einstein equation. Passing to cohomology classes yields the necessary proportionality condition

$$c_1(K_X) = \lambda[\omega] \tag{2-1}$$

in $H^2(X, \mathbb{R})$, which implies that the canonical bundle has a sign: X is either *canonically polarized* ($\lambda > 0$), *Calabi-Yau* ($\lambda = 0$) or *Fano* ($\lambda < 0$).

(c) Finally, a *constant scalar curvature Kähler metric* (cscK for short) is a Kähler metric ω with $S(\omega)$ constant, i.e. $S(\omega) = -n\lambda$. Here the sign of λ only gives very weak information on the positivity properties of K_X. Note that $S(\omega)$ is constant iff the Ricci potential ρ is harmonic, hence constant by compactness of X.

While a KE metric ω is trivially cscK, it is remarkable that the converse is also true as soon as the (necessary) cohomological proportionality condition holds, the reason being

$$(2\text{-}1) \Longrightarrow \mathrm{Ric}(\omega) = -\lambda\omega + dd^c \rho. \tag{2-2}$$

This follows indeed from the $\partial\bar\partial$-lemma, which states that an exact (p, q)-form on a compact Kähler manifold is $\partial\bar\partial$-exact, hence Equation (2-1) $\Longleftrightarrow$ $\mathrm{Ric}(\omega) = -\lambda\omega + dd^c f$ for some $f \in C^\infty(X)$. Taking the trace with respect to ω shows that $f - \rho$ is harmonic, hence constant, proving Equation (2-2).

Thanks to the same $\partial\bar\partial$-lemma, one can introduce *global* potentials for Kähler metrics in a fixed cohomology class. More precisely, given a Kähler form ω, any other Kähler form in the cohomology class of ω is of the form $\omega_u := \omega + dd^c u$ with u a *Kähler potential*, i.e. an element of the open, convex set of smooth functions

$$\mathcal{H} := \{u \in C^\infty(X) \mid \omega_u > 0\}.$$

Assuming Equation (2-1), and hence $\mathrm{Ric}(\omega) = -\lambda\omega + dd^c \rho$, a simple computation yields

$$\mathrm{Ric}(\omega_u) + \lambda\omega_u = dd^c \log\left(\frac{e^{\lambda u}\mu_0}{\omega_u^n}\right),$$

and ω_u is thus Kähler-Einstein iff u satisfies the complex Monge-Ampère equation

$$\mathrm{MA}(u) := V^{-1}\omega_u^n = c\, e^{\lambda u}\mu_0 \tag{2-3}$$

where $c > 0$ is a normalizing constant ensuring that the right-hand side is a probability measure.

2.2 Energy functionals. A fundamental feature of the cscK problem, discovered by Mabuchi [1986], is that the corresponding (fourth order) PDE $S(\omega_u) + n\lambda = 0$ for a potential u can be written as the Euler-Lagrange equation of a functional $M : \mathcal{H} \to \mathbb{R}$, the *Mabuchi K-energy functional*. It is characterized by

$$\frac{d}{dt} M(u_t) = -\int_X \dot{u}_t \left(S(\omega_{u_t}) + n\lambda\right) \mathrm{MA}(u_t)$$

for any smooth path (u_t) in $\mathcal{H}$, and normalized by $M(0) = 0$. Note that $M(u)$ is invariant under translation of a constant, hence only depends on the Kähler metric ω_u. The Chen–Tian formula for M Chen [2000a] and Tian [2000] yields a decomposition

$$M = M_{\mathrm{ent}} + M_{\mathrm{pp}},$$

where the *entropy part*

$$M_{\mathrm{ent}}(u) := \int_X \log\left(\frac{\mathrm{MA}(u)}{\mu_0}\right) \mathrm{MA}(u) \in [0, +\infty)$$

is the relative entropy of the probability measure $\mathrm{MA}(u)$ with respect to the Ricci normalized volume form μ_0, and the *pluripotential part* $M_{\mathrm{pp}}(u)$ is a linear combination of terms of the form $\int_X u\, \omega_u^j \wedge \omega^{n-j}$ and $\int_X u\ \mathrm{Ric}(\omega) \wedge \omega_u^j \wedge \omega^{n-j-1}$.

Assume now that the cohomological proportionality condition $c_1(K_X) = \lambda[\omega]$ holds, so that ω_u is cscK iff u satisfies the complex Monge-Ampère Equation (2-3). Besides the K-energy M, another (simpler) functional also has Equation (2-3) as its Euler–Lagrange equation. Indeed, the complex Monge-Ampère operator $\mathrm{MA}(u)$ is the derivative of a functional $E : \mathcal{H} \to \mathbb{R}$, i.e.

$$\frac{d}{dt} E(u_t) = \int_X \dot{u}_t\, \mathrm{MA}(u_t)$$

The functional E, normalized by $E(0) = 0$, is called the *Monge-Ampère energy* (with strong fluctuations in both notation and terminology across the literature), and is explicitly given by

$$E(u) = \frac{1}{n+1} \sum_{j=0}^{n} V^{-1} \int_X u\, \omega_u^j \wedge \omega^{n-j}. \tag{2-4}$$

It follows that ω_u is cscK (equivalently, KE) iff u is a critical point of the *Ding functional* $D : \mathcal{H} \to \mathbb{R}$, defined as $D := L - E$ with

$$L(u) := \begin{cases} \lambda^{-1} \log\left(\int_X e^{\lambda u} \mu_0\right) & \text{if } \lambda \neq 0 \\ \int_X u\, \mu_0 & \text{if } \lambda = 0. \end{cases}$$

Note that $E(u+c) = E(u) + c$ and $L(u+c) = L(u) + c$ for $c \in \mathbb{R}$, so that $D(u)$, just as $M(u)$, is invariant under translation of u by a constant, and hence only depends on the Kähler form ω_u.

3 The variational approach

This section first describes the L^p-geometry of the space of Kähler potentials, with respect to which the K-energy becomes convex. This is used to relate the coercivity of M, its growth along geodesic rays, and the existence of minimizers.

3.1 The Mabuchi L^2-metric and weak geodesics. As we saw above, cscK metrics are characterized as critical points of the K-energy $M : \mathcal{H} \to \mathbb{R}$. In order to set up a variational approach to the cscK problem, an ideal scenario would thus be that M be convex with respect to the linear structure of $\mathcal{H}$ as an open convex subset of the vector space $C^\infty(X)$, which would in particular imply that cscK metrics correspond to minimizers of M.

While convexity in this sense fails in general, Mabuchi realized in Mabuchi [1987] that M does become convex with respect to a more sophisticated notion of geodesics in $\mathcal{H}$. The infinite dimensional manifold $\mathcal{H}$ is indeed endowed with a natural Riemannian metric, defined at $u \in \mathcal{H}$ as the L^2-scalar product with respect to the volume form $\mathrm{MA}(u) = V^{-1}\omega_u^n$. Mabuchi computed the Levi-Civita connection and curvature of this L^2-metric, and proved that the (Riemannian) Hessian of M is everywhere nonnegative, so that M is convex along (smooth) geodesics in $\mathcal{H}$.

The existence of a geodesic joining two given points in $\mathcal{H}$ thus becomes a pressing issue, and new light was shed on this problem in Semmes [1992] and S. K. Donaldson [1999], with the key observation that the equation for geodesics in $\mathcal{H}$ can be rewritten as a complex Monge-Ampère equation. In terms of the one-to-one correspondence between paths $(u_t)_{t\in I}$ of functions on X parametrized by a open interval $I \subset \mathbb{R}$ and S^1-invariant functions U on the product $X \times \mathbb{D}_I$ of X with the annulus

$$\mathbb{D}_I := \{\tau \in \mathbb{C} \mid -\log|\tau| \in I\}$$

given by setting

$$U(x,\tau) = u_{-\log|\tau|}(x), \tag{3-1}$$

a smooth path $(u_t)_{t\in I}$ in $\mathcal{H}$ is a geodesic iff U satisfies the complex Monge-Ampère equation

$$(\omega + dd^c U)^{n+1} = 0. \tag{3-2}$$

Finding a geodesic $(u_t)_{t\in[0,1]}$ joining two given points $u_0, u_1 \in \mathcal{H}$ thus amounts to solving Equation (3-2) with prescribed boundary data. While uniqueness is a simple matter, existence is much more delicate (and turns out to fail in general), as vanishing of the right-hand side makes this nonlinear elliptic equation degenerate. Since the restriction of the $(1,1)$-form $\omega + dd^c U$ to each slice $X \times \{\tau\}$ is required to be positive, Equation (3-2) imposes that $\omega + dd^c U \geq 0$, which means by definition that U is ω-psh (for plurisubharmonic). Thanks to this observation, geodesics can be approached using pluripotential theory.

Denote by $\mathrm{PSH}(X,\omega)$ the space of ω-psh functions on X, i.e. pointwise limits of decreasing sequences in $\mathcal{H}$, by Błocki and Kołodziej [2007]. Following Berndtsson [2015, §2.2], we define a *subgeodesic* in $\mathrm{PSH}(X,\omega)$ as a family $(u_t)_{t\in I}$ of ω-psh functions whose corresponding function U on $X \times \mathbb{D}_I$ is ω-psh, a condition which implies in particular that $u_t(x)$ is a convex function of t. A *weak geodesic* $(u_t)_{t\in I}$ is a subgeodesic which is *maximal*, i.e. for any compact interval $[a,b] \subset I$ and any subgeodesic $(v_t)_{t\in(a,b)}$,

$$\lim_{t\to a} v_t \leq u_a \text{ and } \lim_{t\to b} v_t \leq u_b \Longrightarrow v_t \leq u_t \text{ for } t \in (a,b).$$

Lemma 3.1. *Darvas [2017] Let $(u_t)_{t\in I}$ be a weak geodesic in $\mathrm{PSH}(X,\omega)$, and pick a compact interval $[a,b] \subset I$. If $u_b - u_a$ is bounded above, then $t \mapsto \sup_X(u_t - u_a)$ is affine on $[a,b]$.*

Proof. After reparametrizing, we assume for ease of notation that $a = 0$ and $b = 1$, and set $m := \sup_X(u_1 - u_0)$. For $t \in [0,1]$, the inequality $\sup_X(u_t - u_0) \leq tm$ follows directly from the convexity of $t \mapsto u_t(x)$. Since $v_t(x) := u_1(x) + (t-1)m$ is a subgeodesic with $v_0 \leq u_0$ and $v_1 \leq u_1$, maximality of (u_t) implies $v_t \leq u_t$ for $t \in [0,1]$, and hence

$$tm = \sup_X(u_1 - u_0) + (t-1)m \leq \sup_X(u_t - u_0).$$

□

Given $u_0, u_1 \in \mathrm{PSH}(X,\omega)$, the weak geodesic $(u_t)_{t\in(0,1)}$ joining them is defined as the usc upper envelope of the family of all subgeodesics $(v_t)_{t\in(0,1)}$ such that $\lim_{t\to 0} v_t \leq u_0$, $\lim_{t\to 1} v_t \leq u_1$ (or $u_t \equiv -\infty$ if no such subgeodesic exists). When u_0, u_1 are bounded, the weak geodesic (u_t) is locally bounded, and a 'balayage' argument shows that the corresponding function U is the unique locally bounded solution to Equation (3-2) in the sense of Bedford and Taylor [1976], with the prescribed boundary data. Even for $u_0, u_1 \in \mathcal{H}$, exemples due to Lempert and Vivas [2013] show that the weak geodesic (u_t) joining them is not C^2 in general, but initial work by Chen [2000b], succesively refined in Błocki [2012] and Chu, Tosatti, and Weinkove [2017], eventually established that U is locally $C^{1,1}$.

3.2 L^p-geometry in the space of Kähler potentials. Just as the Riemannian metric on the space of norms $\mathcal{N}$ can be generalized to a Finsler ℓ^p-metric for any $p \in [1, \infty]$ (cf. Section 1.1), it was noticed by T. Darvas that the Mabuchi L^2-metric on $\mathcal{H}$ admits an immediate generalization to an L^p-Finsler metric, by replacing the L^2-norm with the L^p-norm in the above definition. The associated pseudometric d_p on $\mathcal{H}$ is defined by letting $d_p(u, u')$ be the infimum of the L^p-lengths

$$\int_0^1 \|\dot{u}_t\|_{L^p(\mathrm{MA}(u_t))} dt$$

of all smooth paths $(u_t)_{t \in [0,1]}$ in $\mathcal{H}$ joining u to u'. We trivially have $d_p \le d_{p'}$ for $p \le p'$, but the fact that d_p is actually a metric (i.e. separates distinct points) is a nontrivial result in this infinite dimensional setting, proved in Chen [2000b] for $p = 2$ and in Darvas [2015] for d_1, and hence for all d_p.

The space $\mathcal{H}$ is not complete for any of the metrics d_p, and the description of the completion was completely elucidated in Darvas [ibid.] in terms of pluripotential theory, following an earlier attempt by V. Guedj. The class

$$\mathcal{E} \subset \mathrm{PSH}(X, \omega)$$

of ω-psh functions u with *full Monge-Ampère mass*, introduced by Guedj-Zeriahi in Guedj and Zeriahi [2007] (see also Boucksom, Eyssidieux, Guedj, and Zeriahi [2010]), may be described as the largest class of ω-psh functions on which the Monge-Ampère operator $u \mapsto \mathrm{MA}(u)$ is defined and satisfies:

(i) $\mathrm{MA}(u)$ is a probability measure that puts no mass on pluripolar sets, i.e. sets of the form $\{\psi = -\infty\}$ with ψ ω-psh;

(ii) the operator is continuous along decreasing sequences.

For $p \in [1, \infty]$, the class $\mathcal{E}^p \subset \mathcal{E}$ of ω-psh functions with *finite L^p-energy* is defined as the set of $u \in \mathcal{E}$ that are L^p with respect to $\mathrm{MA}(u)$. For domains in $\mathbb{C}^n$, the analogue of $\mathcal{E}^p$ was first introduced by U. Cegrell in his pioneering work Cegrell [1998].

Example 3.2. *If X is a Riemann surface, a function $u \in \mathrm{PSH}(X, \omega)$ belongs to $\mathcal{E}$ iff the measure $\omega + dd^c u$ puts no mass on polar sets, and u is in $\mathcal{E}^1$ iff it satisfies the classical finite energy condition $\int_X du \wedge d^c u < +\infty$, which means that the gradient of u is in L^2.*

The following results are due to T. Darvas.

Theorem 3.3. *Darvas [2015] The metric d_p admits a unique extension to $\mathcal{E}^p$ that is continuous along decreasing sequences, and $(\mathcal{E}^p, d_p)$ is the completion of $(\mathcal{H}, d_p)$. Further:*

(i) $d_p(u, u')$ *is Lipschitz equivalent to* $\|u - u'\|_{L^p(\mathrm{MA}(u))} + \|u - u'\|_{L^p(\mathrm{MA}(u'))}$;

(ii) *the weak geodesic* $(u_t)_{t\in[0,1]}$ *joining any two* $u_0, u_1 \in \mathcal{E}^p$ *is contained in* $\mathcal{E}^p$*, and is a constant speed geodesic in the metric space* $(\mathcal{E}^p, d_p)$*,* i.e. $d_p(u_t, u_{t'}) = c|t - t'|$ *for some constant* c.

3.3 Energy functionals on $\mathcal{E}^1$. The weakest metric d_1 turns out to be the most relevant one for Kähler geometry, due to its close relationship with the Monge-Ampère energy E. By Berman, Boucksom, Guedj, and Zeriahi [2013] and Darvas [2015], mixed Monge-Ampère integrals of the form

$$\int_X u_0\, \omega_{u_1} \wedge \cdots \wedge \omega_{u_n}$$

with $u_i \in \mathcal{E}^1$ are well-defined, and continuous with respect to the u_i in the d_1-topology. In particular, the Monge-Ampère operator is continuous in this topology, and Equation (2-4) yields a continuous extension of E to $\mathcal{E}^1$, which is proved to be convex on subgeodesics, and *affine* on weak geodesics.

Lemma 3.4. *If* $u, u' \in \mathcal{E}^1$ *satisfy* $u \le u'$*, then* $d_1(u, u') = E(u') - E(u)$.

Proof. By monotone regularization, it is enough to prove this for $u, u' \in \mathcal{H}$. The corresponding weak geodesic $(u_t)_{t\in[0,1]}$ is then $C^{1,1}$, and its L^1-length $\int_{t=0}^1 \int_X |\dot{u}_t|\,\mathrm{MA}(u_t)$ computes $d_1(u, u')$. By Lemma 3.1, $u_t(x)$ is a nondecreasing function of t, hence $\dot{u}_t \ge 0$, which yields

$$d_1(u, u') = \int_0^1 dt \int_X \dot{u}_t\,\mathrm{MA}(u_t) = \int_0^1 \left(\frac{d}{dt} E(u_t)\right) dt = E(u') - E(u).$$

□

When dealing with translation invariant functionals such as M and D, it is useful to introduce the translation invariant functional $J : \mathcal{E}^1 \to \mathbb{R}_+$ defined by

$$J(u) := V^{-1} \int_X u\,\omega^n - E(u),$$

which vanishes iff u is constant and satisfies $J(u) = d_1(u, 0) + O(1)$ on functions normalized by $\sup u = 0$, thanks to Lemma 3.4.

Since the pluripotential part $M_{\mathrm{pp}}(u)$ of the Mabuchi K-energy is a linear combination of integrals of the form $\int_X u\,\omega_u^j \wedge \omega^{n-j}$ and $\int_X u\,\mathrm{Ric}(\omega) \wedge \omega_u^j \wedge \omega^{n-j-1}$, it admits a continuous extension $M_{\text{pp}} : \mathcal{E}^1 \to \mathbb{R}$. As to the entropy part M_{ent}, it extends to a lower semicontinuous functional

$$M_{\text{ent}} : \mathcal{E}^1 \to [0, +\infty],$$

by defining $M_{\text{ent}}(u)$ to be the relative entropy of $\mathrm{MA}(u)$ with respect to μ_0. Finiteness of $M_{\text{ent}}(u)$ is a subtle condition, which amounts to saying that $\mathrm{MA}(u)$ has a density f with respect to Lebesgue measure such that $f \log f$ is integrable.

Theorem 3.5. *Berman, Boucksom, Eyssidieux, Guedj, and Zeriahi [2011], Berman and Berndtsson [2017], Chen, L. Li, and Păun [2016], and Berman, Darvas, and Lu [2017] The extended functionals satisfy the following properties.*

(i) For each $C > 0$, the set of $u \in \mathcal{E}^1$ with $\sup_X u = 0$ and $M_{\text{ent}}(u) \le C$ is compact in the d_1-topology.

(ii) $|M_{\text{pp}}(u)| \le AJ(u) + B$ for some constant $A, B > 0$.

(iii) The functional $M : \mathcal{E}^1 \to (-\infty, +\infty]$ is lower semicontinuous and convex on weak geodesics.

3.4 Variational characterization of cscK metrics. The Mabuchi K-energy M is *coercive* if $M \ge \delta J - C$ on $\mathcal{E}^1$ by for some constants $\delta, C > 0$. By Berman, Darvas, and Lu [2017], it is in fact enough to test this on $\mathcal{H}$. We then have the following basic dichotomy.

Theorem 3.6. *Darvas and He [2017], Darvas and Rubinstein [2017], and Berman, Boucksom, and Jonsson [2015] If the K-energy M is coercive, then it admits a minimizer in $\mathcal{E}^1$. If not, then for any $u \in \mathcal{H}$, there exists a unit speed weak geodesic ray $(u_t)_{t\in[0,+\infty)}$ in $\mathcal{E}^1$ emanating from u, normalized by $\sup_X(u_t - u) = 0$, along which $M(u_t)$ is nonincreasing.*

Proof. Assume that M is coercive, and let $u_j \in \mathcal{E}^1$ be a minimizing sequence, which can be normalized by $\sup u_j = 0$ by translation invariance. Since $M(u_j)$ is bounded above, $J(u_j)$ is bounded, by coercivity, hence so is $|M_{\text{pp}}(u_j)| \le AJ(u_j) + B$. As a result, $M_{\text{ent}}(u_j)$ is also bounded, which means that u_j stays in a compact subset of $\mathcal{E}^1$. After passing to a subsequence, we may thus assume that u_j admits a limit $u \in \mathcal{E}^1$, which is a minimizer of M by lower semicontinuity.

Assume now that M is not coercive, i.e. $M(u_j) \le \delta_j J(u_j) - C_j$ for some sequences $u_j \in \mathcal{E}^1$ with $\sup(u_j - u) = 0$, $\delta_j \to 0$ and $C_j \to +\infty$. We then argue as in Theorem 1.6. Since $M_{\text{ent}}(u_j) \ge 0$ and $M_{\text{pp}}(u_j) \ge -AJ(u_j) - B$, $(A + \delta_j)J(u_j) \ge C_j - B$ tends to ∞, hence so does

$$T_j := d_1(u_j, u) = J(u_j) + O(1).$$

Denote by $(u_{j,t})_{t\in[0,T_j]}$ the weak geodesic connecting u to u_j, parametrized so that $d_1(u_{j,t},u_{j,s}) = |t-s|$, and note that $\sup_X(u_{j,t}-u) = 0$ for all t, by Lemma 3.1. By convexity of M along $(u_{j,t})$, we get

$$\frac{M(u_{j,t})-M(u)}{t} \le \frac{M(u_j)-M(u)}{T_j} \le \delta_j. \tag{3-3}$$

for $j \gg 1$. For each $T > 0$ fixed, $|M_{\mathrm{pp}}(u_{j,t})| \le AJ(u_{j,t}) + B$ is bounded for $t \le T$, hence so is $M_{\mathrm{ent}}(u_{j,t})$, by Equation (3-3). By Theorem 3.5, the 1-Lipschitz maps $t \mapsto u_{j,t}$ thus send each compact subset of $\mathbb{R}_+$ to a fixed compact set in $\mathcal{E}^1$, and Ascoli's theorem shows that $(u_{j,t})$ converges uniformly on compact sets of $\mathbb{R}_+$ to a ray $(u_t)_{t\in\mathbb{R}_+}$ in $\mathcal{E}^1$ (after passing to a subsequence). By local uniform convergence, (u_t) is a weak geodesic, and satisfies $\sup(u_t-u) = 0$ and $d_1(u_t,u_s) = |t-s|$. Further, $M(u_t) \le M(u)$ by Equation (3-3) and lower semicontinuity, which implies that $M(u_t)$ decreases, by convexity. □

Using their key convexity result and a perturbation argument, Berman-Berndtsson proved in Berman and Berndtsson [2017] that cscK metrics in the class $[\omega]$ minimize M, and that the identity component $\mathrm{Aut}^0(X)$ of the group of holomorphic automorphisms acts transitively on these metrics. In Berman, Darvas, and Lu [2016], Berman-Darvas-Lu went further and proved that the existence of *one* cscK metric ω_u implies that any other minimizer of M lies in the $\mathrm{Aut}^0(X)$-orbit of u, and hence is smooth. Using this, we have:

Corollary 3.7. *Darvas and Rubinstein [2017] and Berman, Darvas, and Lu [2016] If* $\mathrm{Aut}^0(X)$ *is trivial and M admits a minimizer $u \in \mathcal{H}$, then M is coercive.*

Proof. By Berman, Darvas, and Lu [2016], u is the unique minimizer of M in $\mathcal{E}^1$, up to a constant. Assume by contradiction that M is not coercice, and let (u_t) be the ray constructed in Theorem 3.6. Since $M(u_t) \le M(u) = \inf M$, u_t must be equal to u up to a constant, and hence $u_t = u$ by normalization, which contradicts $d_1(u_t,u) = t$. □

If a minimizer u of M lies in $\mathcal{H}$, then $u+tf$ is in $\mathcal{H}$ for all test functions $f \in C^\infty(X)$ and $0 < t \ll 1$, hence $M(u+tf) \ge M(u)$, which implies that u is a critical point of M, i.e. ω_u is cscK. This simple perturbation argument cannot be performed for a minimizer in $\mathcal{E}^1$, which is a major remaining difficulty[2] on the analytic side of the cscK problem. In the Kähler-Einstein case, we have however:

Theorem 3.8. *Berman, Boucksom, Guedj, and Zeriahi [2013] and Berman, Boucksom, Eyssidieux, Guedj, and Zeriahi [2011] If the cohomological proportionality condition, Equation (2-1) holds, any mimizer of M in $\mathcal{E}^1$ lies in $\mathcal{H}$, and hence defines a Kähler-Einstein metric.*

[2]This difficulty has recently been overcome by Chen and Cheng [2017, 2018a,b].

Proof. It is not hard to show that a minimizer for M is also a minimizer for the Ding functional $D = L - E$, whose critical points in $\mathcal{H}$ are solutions of the complex Monge-Ampère Equation (2-3). The main step is now to prove that a minimizer $u \in \mathcal{E}^1$ of D satisfies Equation (2-3) in the sense of pluripotential theory, for the complex Monge-Ampère arsenal can then be used to infer ultimately that u is smooth. The projection argument to follow goes back to Aleksandrov in the setting of real Monge-Ampère equations. Given a test function $f \in C^\infty(X)$, the *psh envelope* $P(u+f)$ is defined as the largest ω-psh function dominated by $u+f$. The functional L makes sense on any function u, ω-psh or not, and satisfies $u \le v \implies L(u) \le L(v)$. We thus get for each $t > 0$

$$L(u) - E(u) = D(u) \le D(P(u+tf)) \le L(u+tf) - E(P(u+tf)).$$

The key ingredient is now a differentiability result proved in Berman and Boucksom [2010], which implies that $t \mapsto E(P(u+tf))$ is differentiable at 0, with derivative equal to $\int_X f \, \mathrm{MA}(u)$. This yields indeed

$$\int_X f \, \mathrm{MA}(u) = \lim_{t\to 0_+} \frac{E(P(u+tf)) - E(u)}{t} \le$$
$$\le \lim_{t\to 0_+} \frac{L(u+tf) - L(u)}{t} = \frac{\int_X f\, e^{\lambda u}\mu_0}{\int_X e^{\lambda u}\omega_0},$$

which proves, after replacing f with $-f$, that u is a weak solution of Equation (2-3). □

4 Non-Archimedean Kähler geometry and K-stability

In this final section, we turn to the non-Archimedean aspects of the cscK problem. We reformulate K-stability as a positivity property for the non-Archimedean analogue of the K-energy M, and explain how uniform K-stability implies coercivity, in the Kähler-Einstein case.

4.1 Non-Archimedean pluripotential theory. If X is a complex algebraic variety, we denote by X^{NA} its *Berkovich analytification* (viewed as a topological space) with respect to the *trivial absolute value* $|\cdot|_0$ on $\mathbb{C}$ Berkovich [1990]. When $X = \operatorname{Spec} A$ is affine, with A a finitely generated $\mathbb{C}$-algebra, X^{NA} is defined as the set of all multiplicative seminorms $|\cdot| : A \to \mathbb{R}_+$ compatible with $|\cdot|_0$, endowed with the topology of pointwise convergence. In the general case, X can be covered by finitely many affine open sets X_i, and X^{NA} is defined by gluing together the analytifications X_i^{NA} along their common open subsets $(X_i \cap X_j)^{\mathrm{NA}}$.

Assume from now on that X is projective, equipped with an ample line bundle L. The topological space X^{NA} is then compact (Hausdorff), and can be viewed as a compactification of the space of real-valued valuations $v : \mathbb{C}(X)^* \to \mathbb{R}$ on the function field of X, identifying v with the multiplicative norm $|\cdot| = e^{-v}$. In particular, the trivial valuation on $\mathbb{C}(X)$ defines a special point $0 \in X^{\mathrm{NA}}$, fixed under the natural $\mathbb{R}_+^*$-action $(t, |\cdot|) \mapsto |\cdot|^t$.

In this trivially valued setting, (the analytification of) L comes with a canonical *trivial metric*. Any section $s \in H^0(X, L)$ thus defines a continuous function $|s|_0 : X^{\mathrm{NA}} \to [0, 1]$, the value of $-\log|s|_0$ at a valuation v being equal to that of v on the local function corresponding to s in a trivialization of L at the center of v.

The space $\mathcal{H}^{\mathrm{NA}}$ of *non-Archimedean Kähler potentials* (with respect to L) is defined as the set of continuous functions $\varphi \in C^0(X^{\mathrm{NA}})$ of the form

$$\varphi = \frac{1}{k} \max_i \{\log|s_i|_0 + \lambda_i\}$$

with (s_i) a finite set of sections of $H^0(kL)$ without common zeroes and $\lambda_i \in \mathbb{R}$, those with $\lambda_i \in \mathbb{Q}$ forming $\mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}} \subset \mathcal{H}^{\mathrm{NA}}$. In order to motivate this definition, recall that the data of a Hermitian norm γ on $H^0(kL)$ defines a Fubini-Study/Bergman type metric on L, whose potential with respect to a reference metric $|\cdot|_0$ on L can be written as

$$\mathrm{FS}_k(\gamma) := \frac{1}{k} \log \max_{s \in H^0(kL) \setminus \{0\}} \frac{|s|_0}{\gamma(s)} = \frac{1}{2k} \log \sum_i |s_i|_0^2$$

for any γ-orthonormal basis (s_i). Similarly, any non-Archimedean norm α on $H^0(kL)$ in the sense of Section 1.2 admits an orthogonal basis (s_i), and we then have

$$\mathrm{FS}_k^{\mathrm{NA}}(\alpha) := \frac{1}{k} \log \max_{s \in H^0(kL) \setminus \{0\}} \frac{|s|_0}{\alpha(s)} = \frac{1}{k} \max_i \{\log|s_i|_0 + \lambda_i\}.$$

with $\lambda_i = -\log \alpha(s_i)$. Denoting respectively by $\mathcal{N}_k$ and $\mathcal{N}_k^{\mathrm{NA}}$ the spaces of Hermitian and non-Archimedean norms on $H^0(kL)$, we thus have two natural maps

$$\mathrm{FS}_k : \mathcal{N}_k \to \mathcal{H}, \quad \mathrm{FS}_k^{\mathrm{NA}} : \mathcal{N}_k^{\mathrm{NA}} \to \mathcal{H}^{\mathrm{NA}},$$

and $\mathcal{H}^{\mathrm{NA}} = \bigcup_k \mathrm{FS}_k^{\mathrm{NA}}(\mathcal{N}_k^{\mathrm{NA}})$ by definition. This is to be compared with the fact that $\bigcup_k \mathrm{FS}_k(\mathcal{N}_k)$ is dense in $\mathcal{H}$, a consequence of the fundamental Bouche–Catlin–Tian–Zelditch asymptotic expansion of Bergman kernels.

Non-Archimedean Kähler potentials are closely related to *test configurations* for (X, L), i.e. $\mathbb{C}^*$-equivariant partial compactifications $(\mathcal{X}, \mathcal{L}) \to \mathbb{C}$ of the product $(X, L) \times \mathbb{C}^*$, with $\mathcal{L}$ a $\mathbb{Q}$-line bundle.

Proposition 4.1. *Every test configuration* $(\mathcal{X}, \mathcal{L})$ *gives rise in a natural way to a function* $\varphi_{\mathcal{L}} \in C^0(X^{\mathrm{NA}})$*, which belongs to* $\mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}}$ *if* $\mathcal{L}$ *is ample, and is a difference of functions in* $\mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}}$ *in general. Further, two test configurations* $(\mathcal{X}_i, \mathcal{L}_i)$*,* $i = 1, 2$ *yield the same function on* X^{NA} *if and only if* $\mathcal{L}_1$ *and* $\mathcal{L}_2$ *coincide after pulling-back to some higher test configuration.*

To define $\varphi_{\mathcal{L}}$, denote respectively by $\mathcal{L}'$ and $L_{\mathcal{X}'}$ the pullbacks of $\mathcal{L}$ and L to the graph $\mathcal{X}'$ of the canonical $\mathbb{C}^*$-equivariant birational map $\mathcal{X} \dashrightarrow X \times \mathbb{C}$, and note that

$$\mathcal{L}' = L_{\mathcal{X}'} + D$$

for a unique $\mathbb{Q}$-Cartier divisor D supported in the central fiber $\mathcal{X}'_0$. Every valuation v on X admits a natural $\mathbb{C}^*$-invariant (Gauss) extension $G(v)$ to $\mathbb{C}(X)(t) \simeq \mathbb{C}(\mathcal{X}')$, which can be evaluated on D by chosing a local equation for (a Cartier multiple of) D at the center of $G(v)$, and we set $\varphi_{\mathcal{L}}(v) := G(v)(D)$.

Example 4.2. *Every* 1*-parameter subgroup* $\rho : \mathbb{C}^* \to \mathrm{GL}(H^0(kL))$ *with* kL *very ample defines a test configuration* $(\mathcal{X}, \mathcal{L})$*, obtained as the closure of the orbit of* $X \hookrightarrow \mathbb{P}H^0(kL)^*$*. The* $\mathbb{Q}$*-line bundle* $\mathcal{L}$ *is ample, and every test configuration* $(\mathcal{X}, \mathcal{L})$ *with* $\mathcal{L}$ *ample arises this way. By Example 1.5,* ρ *also defines a non-Archimedean norm* α_ρ *on* $H^0(kL)$*, and we have*

$$\varphi_{\mathcal{L}} = \mathrm{FS}^{\mathrm{NA}}_k(\alpha_\rho).$$

Combined with Proposition 4.1, this implies that $\mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}}$ *is in one-to-one correspondence with the set of all normal, ample test configurations.*

A more general *L-psh function* is defined as a usc function $\varphi : X^{\mathrm{NA}} \to [-\infty, +\infty)$ that can be written as the pointwise limit of a decreasing sequence (or net, rather) in $\mathcal{H}^{\mathrm{NA}}$, defining a space $\mathrm{PSH}^{\mathrm{NA}}$. These functions are bounded above, and the maximum principle takes the simple form

$$\sup_{X^{\mathrm{NA}}} \varphi = \varphi(0),$$

with $0 \in X^{\mathrm{NA}}$ the trivial valuation. The space $\mathrm{PSH}^{\mathrm{NA}}$ is endowed with a natural topology of pointwise convergence on divisorial points, in which functions with $\sup \varphi = 0$ form a compact set. This is proved in Boucksom and Jonsson [2018], building on previous work Boucksom, Favre, and Jonsson [2016] dealing with Berkovich spaces over the field $\mathbb{C}((t))$ of formal Laurent series.

Example 4.3. *If* $\mathfrak{a}$ *is a coherent ideal sheaf on* X*, setting* $|\mathfrak{a}| = \max_{f \in \mathfrak{a}} |f|$ *defines a continuous function* $|\mathfrak{a}| : X^{\mathrm{NA}} \to [0, 1]$*. Given* $c > 0$*, one shows that the function* $c \log |\mathfrak{a}|$ *is* L*-psh if and only if* $L \otimes \mathfrak{a}^c$ *is nef, in the sense that* $\mu^* L - cE$ *is nef on the normalized blow-up* $\mu : X' \to X$ *of* $\mathfrak{a}$*, with exceptional divisor* E*.*

4.2 From geodesic rays to non-Archimedean potentials. We assume from now on that X is a projective manifold equipped with an ample line bundle L, and $\omega \in c_1(L)$ is a Kähler form. Recall that a subgeodesic ray $(u_t)_{t\in\mathbb{R}_+}$ in $\mathrm{PSH}(X,\omega)$ is encoded in the associated S^1-invariant ω-psh function

$$U(x,\tau) = u_{-\log|\tau|}(x)$$

on $X \times \mathbb{D}^*$. We shall say that (u_t) has *linear growth* if $u_t \le at + b$ for some constants $a, b \in \mathbb{R}$, i.e. $U + a\log|\tau| \le b$, a condition which automatically holds when (u_t) is a weak geodesic ray emanating from $u_0 \in \mathcal{H}$, as a consequence of Lemma 3.1.

To a subgeodesic ray (u_t) with linear growth, we shall associate an L-psh function

$$U^{\mathrm{NA}} : X^{\mathrm{NA}} \to [-\infty, +\infty),$$

following a procedure initiated in Boucksom, Favre, and Jonsson [2008]. Imposing

$$(U + a\log|\tau|)^{\mathrm{NA}} = U^{\mathrm{NA}} - a,$$

we may assume that U is bounded above, and hence extends to an ω-psh function on $X \times \mathbb{D}$. Consider first the case where U has *analytic singularities*, i.e. locally satisfies

$$U = c\log\max_i |f_i| + O(1)$$

for a fixed constant $c > 0$ and finitely many holomorphic functions (f_i). The (integrally closed) ideal sheaf

$$\mathfrak{a} := \{f \in \mathcal{O}_{X\times\mathbb{D}} \mid c\log|f| \le U + O(1)\}$$

is then coherent, and $\mathbb{C}^*$-invariant by S^1-invariance of U. We thus have a weight decomposition $\mathfrak{a} = \sum_{i=0}^{r} \tau^i \mathfrak{a}_i$ for an increasing sequence of coherent ideal sheaves $\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \cdots \subset \mathfrak{a}_r = \mathcal{O}_X$ on X. One further proves that $L \otimes \mathfrak{a}_i^c$ is nef for each i, yielding L-psh functions $c\log|\mathfrak{a}_i|$ on X^{NA} by Example 4.3, and hence an L-psh function

$$U^{\mathrm{NA}} = c\log|\mathfrak{a}| := c\max_i\{\log|\mathfrak{a}_i| - i\}.$$

In the general case, the *multiplier ideals* $\mathcal{J}(kU)$, $k \in \mathbb{N}^*$, are $\mathbb{C}^*$-invariant coherent ideal sheaves on $X \times \mathbb{D}$. They satisfy the fundamental subadditivity property

$$\mathcal{J}((k+k')U) \subset \mathcal{J}(kU)\cdot\mathcal{J}(k'U),$$

which implies the existence of the pointwise limit

$$U^{\mathrm{NA}} := \lim_{k\to\infty} \tfrac{1}{k}\log|\mathcal{J}(kU)|$$

on X^{NA}. A variant of Siu's uniform generation theorem Berman, Boucksom, and Jonsson [2015, §3.2] further shows the existence of k_0 such that $p_X^*((k+k_0)L) \otimes \mathcal{J}(kU)$ is globally generated on $X \times \mathbb{D}$ for all k, and it follows that U^{NA} is indeed L-psh.

Example 4.4. *Pick $k \gg 1$, and let $\gamma_t = \iota_s(t\lambda)$ be a geodesic ray in $\mathcal{N}_k$ associated to a basis $s = (s_i)$ of $H^0(kL)$ and $\lambda \in \mathbb{R}^N$. The image $u_t := \mathrm{FS}_k(\gamma_t)$ is then a subgeodesic ray in $\mathcal{H}$ with linear growth, and $U^{\mathrm{NA}} = \mathrm{FS}_k^{\mathrm{NA}}\left(\iota_s^{\mathrm{NA}}(\lambda)\right)$ is the image of the non-Archimedean norm defined by (γ_t). If λ is further rational, U has analytic singularities, and blowing-up $X \times \mathbb{C}$ along the associated $\mathbb{C}^*$-invariant ideal $\mathfrak{a}$ defines a test configuration $(\mathcal{X}, \mathcal{L})$ such that $U^{\mathrm{NA}} = \varphi_{\mathcal{L}}$.*

The function U^{NA} basically captures the Lelong numbers of U, and we have in particular $U^{\mathrm{NA}} = 0$ iff U has zero Lelong numbers at all points of $X \times \{0\}$. More specifically, let $\mathcal{X}$ be a normal test configuration for X, pick an irreducible component E of the central fiber $\mathcal{X}_0$, and set $b_E := \mathrm{ord}_E(\mathcal{X}_0)$. The normalized $\mathbb{C}^*$-invariant valuation $b_E^{-1}\,\mathrm{ord}_E$ on $\mathbb{C}(\mathcal{X}) \simeq \mathbb{C}(X)(\tau)$ restricts to a divisorial (or trivial) valuation on $\mathbb{C}(X)$, defining a point $x_E \in X^{\mathrm{NA}}$. By Boucksom, Hisamoto, and Jonsson [2017, Theorem 4.6], every divisorial point in X^{NA} is of this type, which means that U^{NA} is determined by its values on such points, and Demailly's work on multiplier ideals shows that

$$-b_E U^{\mathrm{NA}}(x_E) = \lim_{k\to\infty} \tfrac{1}{k}\,\mathrm{ord}_E\left(\mathcal{J}(kU)\right)$$

coincides with the generic Lelong number along E of the pull-back of U (cf. Boucksom, Favre, and Jonsson [2008, Proposition 5.6]).

4.3 Non-Archimedean energy functionals. Ideally, we would like to associate to each functional F in Section 2.2 a non-Archimedean analogue F^{NA}, in such a way that

$$F^{\mathrm{NA}}(U^{\mathrm{NA}}) = \lim_{t\to\infty} \frac{F(u_t)}{t} \tag{4-1}$$

for all weak geodesic rays (u_t). To get started, a special case of the pioneering work of A.Chambert-Loir and A.Ducros on forms and currents in Berkovich geometry Chambert-Loir and Ducros [2012] enables to define a mixed non-Archimedean Monge-Ampère operator

$$(\varphi_1, \dots, \varphi_n) \mapsto \mathrm{MA}(\varphi_1, \dots, \varphi_n) \tag{4-2}$$

on n-tuples (φ_i) in $\mathcal{H}^{\mathrm{NA}}$, with values in *atomic* probability measures on X^{NA}. When the φ_i arise from test configurations $(\mathcal{X}_i, \mathcal{L}_i)$, we can assume after pulling back that all $\mathcal{X}_i$ are equal to the same $\mathcal{X}$, and we then have

$$\mathrm{MA}(\varphi_1, \dots, \varphi_n) = \sum_E b_E (\mathcal{L}_1|_E \cdot \ldots \cdot \mathcal{L}_n|_E)\delta_{x_E},$$

where $\mathcal{X}_0 = \sum_E b_E E$ is the irreducible decomposition and the $x_E \in X^{\mathrm{NA}}$ are the associated divisorial points.

We next introduce the *non-Archimedean Monge-Ampère energy* $E^{\mathrm{NA}} : \mathcal{H}^{\mathrm{NA}} \to \mathbb{R}$ using the analogue of Equation (2-4). As in the Kähler case, E^{NA} is nondecreasing, hence extends by monotonicity to $\mathrm{PSH}^{\mathrm{NA}}$, which defines a space

$$\mathcal{E}^{1,\mathrm{NA}} := \{E^{\mathrm{NA}} > -\infty\} \subset \mathrm{PSH}^{\mathrm{NA}}$$

of *L-psh functions φ with finite L^1-energy*. It is proved in Boucksom, Favre, and Jonsson [2015] and Boucksom and Jonsson [2018] that the mixed Monge-Ampère operator Equation (4-2) admits a unique extension to $\mathcal{E}^{1,\mathrm{NA}}$ with the usual continuity property along monotonic sequences, and that

$$J^{\mathrm{NA}}(\varphi) := \sup \varphi - E^{\mathrm{NA}}(\varphi) \in [0, +\infty)$$

vanishes iff $\varphi \in \mathcal{E}^1$ is constant.

Example 4.5. *A test configuration $(\mathcal{X}, \mathcal{L}) \to \mathbb{C}$, being a product away from the central fiber, admits a natural compactification $(\bar{\mathcal{X}}, \bar{\mathcal{L}}) \to \mathbb{P}^1$. The non-Archimedean Monge-Ampère energy $E^{\mathrm{NA}}(\varphi)$ of the corresponding function $\varphi = \varphi_{\mathcal{L}} \in \mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}}$ is then equal to the self-intersection number $\left(c_1(\bar{\mathcal{L}})^{n+1}\right)$, up to a normalization factor. Alternatively,*

$$E^{\mathrm{NA}}(\varphi) = \lim_{k\to\infty} \frac{w_k}{k h^0(kL)}$$

with $w_k \in \mathbb{Z}$ the weight of the induced $\mathbb{C}^*$-action on the determinant line $\det H^0(\mathcal{X}_0, k\mathcal{L}_0)$*, see for instance Boucksom, Hisamoto, and Jonsson [2017, §7.1].*

If (u_t) is a weak geodesic ray in $\mathcal{E}^1$, $E(u_t) = at + b$ is affine. Using that U is more singular than $\mathcal{J}(kU)^{1/k}$, one shows that

$$E^{\mathrm{NA}}(U^{\mathrm{NA}}) \geq a = \lim_{t\to\infty} E(u_t)/t, \tag{4-3}$$

which implies in particular that U^{NA} belongs to $\mathcal{E}^{1,\mathrm{NA}}$. However, this inequality can be strict in general without further assumptions.

Example 4.6. *Let ω be the Fubini-Study metric on $X = \mathbb{P}^1$, normalized to mass 1. A compact, polar Cantor set $K \subset \mathbb{P}^1$ carries a natural probability measure without atoms, and the potential u of this measure with respect to ω is smooth outside K, has zero Lelong numbers and does not belong to $\mathcal{E}$. By Ross and Witt Nyström [2014] and Darvas [2017], u defines a locally bounded weak geodesic ray (u_t) emanating from 0 such that $E(u_t) = at$ with $a < 0$. However, the corresponding ω-psh function U on $X \times \mathbb{D}$ has zero Lelong numbers, hence $U^{\mathrm{NA}} = 0$ and $E^{\mathrm{NA}}(U^{\mathrm{NA}}) = 0$.*

The Mabuchi K-energy M and the Ding functional D also admit non-Archimedean analogues M^{NA} and D^{NA}. While the pluripotential part $M^{\mathrm{NA}}_{\mathrm{pp}}$ of M^{NA} is defined in complete analogy with M_{pp} as a linear combination of mixed Monge-Ampère integrals, the entropy part $M^{\mathrm{NA}}_{\mathrm{ent}}$ as well as L^{NA} turn out to be of a completely different nature, involving the *log discrepancy function*

$$A_X : X^{\mathrm{NA}} \to [0, +\infty].$$

The latter is the maximal lower semicontinuous extension of the usual log discrepancy on divisorial valuations, and we then have

$$M^{\mathrm{NA}}_{\mathrm{pp}}(\varphi) = \int_{X^{\mathrm{NA}}} A_X \,\mathrm{MA}(\varphi)$$

and

$$L^{\mathrm{NA}}(\varphi) = \begin{cases} \lambda^{-1} \inf_{X^{\mathrm{NA}}}(A_X + \lambda\varphi) & \text{if } \lambda \neq 0 \\ \sup_{X^{\mathrm{NA}}} \varphi = \varphi(0) & \text{if } \lambda = 0. \end{cases}$$

where we have set as before $\lambda = V^{-1}(K_X \cdot L^{n-1})$.

Example 4.7. *Boucksom, Hisamoto, and Jonsson [2017] If $(\mathcal{X}, \mathcal{L})$ is an ample test configuration, then $M^{\mathrm{NA}}(\varphi)$ coincides with the* Donaldson-Futaki *invariant of $(\mathcal{X}, \mathcal{L})$, up to a nonnegative error term that vanishes precisely when $\mathcal{X}_0$ is reduced. Further, (X, L) is* K-semistable *iff $M^{\mathrm{NA}}(\varphi) \geq 0$ for all $\varphi \in \mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}}$, and* K-stable *iff equality holds only for φ a constant. Following Boucksom, Hisamoto, and Jonsson [2017] and Dervan [2016], we say that (X, L) is* uniformly K-stable *if $M^{\mathrm{NA}} \geq \delta J^{\mathrm{NA}}$ on $\mathcal{H}^{\mathrm{NA}}_{\mathbb{Q}}$ for some $\delta > 0$.*

We can now state the following result, which builds in part on previous work by Phong, Ross, and Sturm [2008] and Berman [2016].

Theorem 4.8. *Boucksom, Hisamoto, and Jonsson [2016] and Berman, Boucksom, and Jonsson [2015] Let (u_t) be any subgeodesic ray in $\mathcal{E}^1$, normalized by* $\sup u_t = 0$.

(i) If (u_t) has analytic singularities, then Equation (4-1) holds for E and M_{pp}.

(ii) If (u_t) has strongly analytic singularities, then Equation (4-1) holds for M_{ent}.

(iii) In the Fano case, Equation (4-1) holds for L.

Here we say that (u_t) (or U) has *strongly analytic singularities* if U satisfies near each point of $X \times \{0\}$

$$U = \tfrac{c}{2} \log \sum_i |f_i|^2 \bmod C^\infty$$

for a fixed constant $c > 0$ and finitely many holomorphic functions (f_i).

4.4 A version of the Yau–Tian–Donaldson conjecture. In its usual formulation, the Yau–Tian–Donaldson conjecture states that $c_1(L)$ contains a cscK metric if and only if (X, L) is K-(poly)stable. In the following form, it says that M satisfies the analogue of Theorem 1.6.

Conjecture 4.9. *Let (X, L) be a polarized projective manifold, $\omega \in c_1(L)$ be a Kähler form, and assume that $\mathrm{Aut}^0(X, L) = \mathbb{C}^*$. The following are equivalent:*

(i) there exists a cscK metric in $c_1(L)$;

(ii) M is coercive;

(iii) (X, L) is uniformly K-stable.

The implications (i)$\Longrightarrow$(ii)$\Longrightarrow$(iii) were respectively proved in Berman, Darvas, and Lu [2016] (cf. Corollary 3.7) and Boucksom, Hisamoto, and Jonsson [2016] (cf. Theorem 4.8). By Theorem 3.6, (ii) implies the existence of a minimizer $u \in \mathcal{E}^1$ for M, and the key obstacle to get (i) is then to establish that u is smooth[3]. Assume now that (iii) holds. If (ii) fails, Theorem 3.6 yields a weak geodesic ray (u_t) in $\mathcal{E}^1$, emanating from 0 and normalized by $\sup u_t = 0$, $E(u_t) = -t$, along which $M(u_t)$ decreases, and hence $\lim M(u_t)/t \le 0$. Two major difficulties arise:

1. While U^{NA} belongs to $\mathcal{E}^{1,\mathrm{NA}}$, we cannot prove at the moment of this writing that (iii) propagates to $M^{\mathrm{NA}} \ge \delta J^{\mathrm{NA}}$ on the whole of $\mathcal{E}^{1,\mathrm{NA}}$.
2. Even taking (1) for granted, Example 4.6 shows that $M^{\mathrm{NA}}(U^{\mathrm{NA}})$ cannot be expected to compute exactly the slope at infinity of $M(u_t)$.

These difficulties can be overcome in the Kähler-Einstein case, by relying on the Ding functional as well.

Theorem 4.10. *Berman, Boucksom, and Jonsson [2015] Conjecture 4.9 holds if the proportionality condition $c_1(K_X) = \lambda[\omega]$ is satisfied.*

Sketch of proof. For $\lambda \ge 0$, all three conditions in the conjecture are known to be always satisfied, and we thus focus on the Fano case. Theorem 3.8 completes the proof of (ii)$\Longrightarrow$(i), which was anyway proved long before Tian [2000] by using Aubin's continuity method. Assume (iii), and consider a ray (u_t) as above. In the Fano case, we have $M \ge D$, which shows that $D(u_t) = L(u_t) - E(u_t)$ is bounded above as well. We infer from Theorem 4.8 that $\varphi := U^{\mathrm{NA}}$ satisfies

$$L^{\mathrm{NA}}(\varphi) = \lim_{t\to\infty} \frac{L(u_t)}{t} \le \lim_{t\to\infty} \frac{E(u_t)}{t} = -1 \le E^{\mathrm{NA}}(\varphi).$$

[3] As already mentioned, this has recently been overcome by Chen and Cheng [2017, 2018a,b]

Relying on the Minimal Model Program along the same lines as C. Li and Xu [2014], one proves on the other hand that (iii) implies $D^{\mathrm{NA}} \geq \delta J^{\mathrm{NA}}$ on $\mathcal{H}^{\mathrm{NA}}$, and then on $\mathcal{E}^{1,\mathrm{NA}}$ as well. As φ is normalized by $\sup \varphi = 0$, this means

$$L^{\mathrm{NA}}(\varphi) \geq (1-\delta) E^{\mathrm{NA}}(\varphi) \geq \delta - 1,$$

a contradiction. □

References

Eric Bedford and B. A. Taylor (1976). "The Dirichlet problem for a complex Monge-Ampère equation". *Invent. Math.* 37.1, pp. 1–44. MR: 0445006 (cit. on p. 621).

Vladimir G. Berkovich (1990). *Spectral theory and analytic geometry over non-Archimedean fields*. Vol. 33. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. x+169. MR: 1070709 (cit. on p. 626).

Robert J. Berman (2016). "K-polystability of $\mathbb{Q}$-Fano varieties admitting Kähler-Einstein metrics". *Invent. Math.* 203.3, pp. 973–1025. MR: 3461370 (cit. on p. 632).

Robert J. Berman and Bo Berndtsson (2017). "Convexity of the K-energy on the space of Kähler metrics and uniqueness of extremal metrics". *J. Amer. Math. Soc.* 30.4, pp. 1165–1196. MR: 3671939 (cit. on pp. 611, 624, 625).

Robert J. Berman and Sébastien Boucksom (2010). "Growth of balls of holomorphic sections and energy at equilibrium". *Invent. Math.* 181.2, pp. 337–394. MR: 2657428 (cit. on p. 626).

Robert J. Berman, Sébastien Boucksom, Philippe Eyssidieux, Vincent Guedj, and Ahmed Zeriahi (Nov. 2011). "Kähler-Einstein metrics and the Kähler-Ricci flow on log Fano varieties". To appear in J. Reine Angew. Math. arXiv: 1111.7158 (cit. on pp. 611, 624, 625).

Robert J. Berman, Sébastien Boucksom, Vincent Guedj, and Ahmed Zeriahi (2013). "A variational approach to complex Monge-Ampère equations". *Publ. Math. Inst. Hautes Études Sci.* 117, pp. 179–245. MR: 3090260 (cit. on pp. 623, 625).

Robert J. Berman, Sébastien Boucksom, and Mattias Jonsson (Sept. 2015). "A variational approach to the Yau–Tian–Donaldson conjecture". arXiv: 1509.04561 (cit. on pp. 611, 624, 630, 632, 633).

Robert J. Berman, Tamás Darvas, and Chinh H. Lu (Feb. 2016). "Regularity of weak minimizers of the K-energy and applications to properness and K-stability". arXiv: 1602.03114 (cit. on pp. 625, 633).

– (2017). "Convexity of the extended K-energy and the large time behavior of the weak Calabi flow". *Geom. Topol.* 21.5, pp. 2945–2988. MR: 3687111 (cit. on p. 624).

Bo Berndtsson (2015). "A Brunn-Minkowski type inequality for Fano manifolds and some uniqueness theorems in Kähler geometry". *Invent. Math.* 200.1, pp. 149–200. MR: 3323577 (cit. on p. 621).

Zbigniew Błocki (2012). "On geodesics in the space of Kähler metrics". In: *Advances in geometric analysis*. Ed. by Stanislaw Janeczko, Jun Li, and Duong H. Phong. Vol. 21. Adv. Lect. Math. (ALM). Int. Press, Somerville, MA, pp. 3–19. MR: 3077245 (cit. on p. 621).

Zbigniew Błocki and Sławomir Kołodziej (2007). "On regularization of plurisubharmonic functions on manifolds". *Proc. Amer. Math. Soc.* 135.7, pp. 2089–2093. MR: 2299485 (cit. on p. 621).

Sébastien Boucksom and Dennis Eriksson (May 2018). "Spaces of norms, determinant of cohomology and Fekete points in non-Archimedean geometry". arXiv: 1805.01016 (cit. on pp. 613, 614).

Sébastien Boucksom, Philippe Eyssidieux, Vincent Guedj, and Ahmed Zeriahi (2010). "Monge-Ampère equations in big cohomology classes". *Acta Math.* 205.2, pp. 199–262. MR: 2746347 (cit. on p. 622).

Sébastien Boucksom, Charles Favre, and Mattias Jonsson (2008). "Valuations and plurisubharmonic singularities". *Publ. Res. Inst. Math. Sci.* 44.2, pp. 449–494. MR: 2426355 (cit. on pp. 629, 630).

– (2015). "Solution to a non-Archimedean Monge-Ampère equation". *J. Amer. Math. Soc.* 28.3, pp. 617–667. MR: 3327532 (cit. on pp. 611, 631).

– (2016). "Singular semipositive metrics in non-Archimedean geometry". *J. Algebraic Geom.* 25.1, pp. 77–139. MR: 3419957 (cit. on pp. 611, 628).

Sébastien Boucksom, Tomoyuki Hisamoto, and Mattias Jonsson (Mar. 2016). "Uniform K-stability and asymptotics of energy functionals in Kähler geometry". To appear in J. Eur. Math. Soc. arXiv: 1603.01026 (cit. on pp. 632, 633).

– (2017). "Uniform K-stability, Duistermaat-Heckman measures and singularities of pairs". *Ann. Inst. Fourier (Grenoble)* 67.2, pp. 743–841. MR: 3669511 (cit. on pp. 630–632).

Sébastien Boucksom and Mattias Jonsson (Jan. 2018). "Singular semipositive metrics on line bundles on varieties over trivially valued fields". arXiv: 1801.08229 (cit. on pp. 628, 631).

Urban Cegrell (1998). "Pluricomplex energy". *Acta Math.* 180.2, pp. 187–217. MR: 1638768 (cit. on p. 622).

Antoine Chambert-Loir and Antoine Ducros (Apr. 2012). "Formes différentielles réelles et courants sur les espaces de Berkovich". arXiv: 1204.6277 (cit. on p. 630).

Xiuxiong Chen (2000a). "On the lower bound of the Mabuchi energy and its application". *Internat. Math. Res. Notices* 12, pp. 607–623. MR: 1772078 (cit. on p. 619).

– (2000b). "The space of Kähler metrics". *J. Differential Geom.* 56.2, pp. 189–234. MR: 1863016 (cit. on pp. 611, 621, 622).

Xiuxiong Chen and Jingrui Cheng (Dec. 2017). "On the constant scalar curvature Kähler metrics, apriori estimates". arXiv: 1712.06697 (cit. on pp. 611, 625, 633).
– (Jan. 2018a). "On the constant scalar curvature Kähler metrics, existence results". arXiv: 1801.00656 (cit. on pp. 611, 625, 633).
– (Jan. 2018b). "On the constant scalar curvature Kähler metrics, general automorphism group". arXiv: 1801.05907 (cit. on pp. 611, 625, 633).
Xiuxiong Chen, Simon Donaldson, and Song Sun (2015a). "Kähler-Einstein metrics on Fano manifolds. I: Approximation of metrics with cone singularities". *J. Amer. Math. Soc.* 28.1, pp. 183–197. MR: 3264766 (cit. on p. 610).
– (2015b). "Kähler-Einstein metrics on Fano manifolds. II: Limits with cone angle less than 2π". *J. Amer. Math. Soc.* 28.1, pp. 199–234. MR: 3264767 (cit. on p. 610).
– (2015c). "Kähler-Einstein metrics on Fano manifolds. III: Limits as cone angle approaches 2π and completion of the main proof". *J. Amer. Math. Soc.* 28.1, pp. 235–278. MR: 3264768 (cit. on p. 610).
Xiuxiong Chen, Long Li, and Mihai Păun (2016). "Approximation of weak geodesics and subharmonicity of Mabuchi energy". *Ann. Fac. Sci. Toulouse Math. (6)* 25.5, pp. 935–957. MR: 3582114 (cit. on p. 624).
Jianchun Chu, Valentino Tosatti, and Ben Weinkove (2017). "On the $C^{1,1}$ regularity of geodesics in the space of Kähler metrics". *Ann. PDE* 3.2, Art. 15, 12. MR: 3695402 (cit. on p. 621).
Tamás Darvas (2015). "The Mabuchi geometry of finite energy classes". *Adv. Math.* 285, pp. 182–219. MR: 3406499 (cit. on pp. 611, 622, 623).
– (2017). "Weak geodesic rays in the space of Kähler potentials and the class $\mathcal{E}(X,\omega)$". *J. Inst. Math. Jussieu* 16.4, pp. 837–858. MR: 3680345 (cit. on pp. 621, 631).
Tamás Darvas and Weiyong He (2017). "Geodesic rays and Kähler-Ricci trajectories on Fano manifolds". *Trans. Amer. Math. Soc.* 369.7, pp. 5069–5085. MR: 3632560 (cit. on p. 624).
Tamás Darvas and Yanir A. Rubinstein (2017). "Tian's properness conjectures and Finsler geometry of the space of Kähler metrics". *J. Amer. Math. Soc.* 30.2, pp. 347–387. MR: 3600039 (cit. on pp. 611, 624, 625).
Ruadhaí Dervan (2016). "Uniform stability of twisted constant scalar curvature Kähler metrics". *Int. Math. Res. Not. IMRN* 15, pp. 4728–4783. MR: 3564626 (cit. on p. 632).
Simon K. Donaldson (1999). "Symmetric spaces, Kähler geometry and Hamiltonian dynamics". In: *Northern California Symplectic Geometry Seminar*. Vol. 196. Amer. Math. Soc. Transl. Ser. 2. Amer. Math. Soc., Providence, RI, pp. 13–33. MR: 1736211 (cit. on p. 620).
Vincent Guedj and Ahmed Zeriahi (2007). "The weighted Monge-Ampère energy of quasiplurisubharmonic functions". *J. Funct. Anal.* 250.2, pp. 442–482. MR: 2352488 (cit. on p. 622).

Erich Kähler (1933). "Über eine bemerkenswerte Hermitesche Metrik". *Abh. Math. Sem. Univ. Hamburg* 9.1, pp. 173–186. MR: 3069594 (cit. on p. 610).

Michael Kapovich, Bernhard Leeb, and John Millson (2009). "Convex functions on symmetric spaces, side lengths of polygons and the stability inequalities for weighted configurations at infinity". *J. Differential Geom.* 81.2, pp. 297–354. MR: 2472176 (cit. on p. 612).

Maxim Kontsevich and Yuri Tschinkel (2000). "Non-Archimedean Kähler geometry" (cit. on p. 611).

László Lempert and Liz Vivas (2013). "Geodesics in the space of Kähler metrics". *Duke Math. J.* 162.7, pp. 1369–1381. MR: 3079251 (cit. on p. 621).

Chi Li and Chenyang Xu (2014). "Special test configuration and K-stability of Fano varieties". *Ann. of Math. (2)* 180.1, pp. 197–232. MR: 3194814 (cit. on p. 634).

Toshiki Mabuchi (1986). "K-energy maps integrating Futaki invariants". *Tohoku Math. J.* 38.4, pp. 575–593 (cit. on p. 619).

– (1987). "Some symplectic geometry on compact Kähler manifolds. I". *Osaka J. Math.* 24.2, pp. 227–252. MR: 909015 (cit. on p. 620).

D. H. Phong, Julius Ross, and Jacob Sturm (2008). "Deligne pairings and the Knudsen-Mumford expansion". *J. Differential Geom.* 78.3, pp. 475–496. MR: 2396251 (cit. on p. 632).

Julius Ross and David Witt Nyström (2014). "Analytic test configurations and geodesic rays". *J. Symplectic Geom.* 12.1, pp. 125–169. MR: 3194078 (cit. on p. 631).

Stephen Semmes (1992). "Complex Monge-Ampère and symplectic manifolds". *Amer. J. Math.* 114.3, pp. 495–550. MR: 1165352 (cit. on p. 620).

Gang Tian (2000). *Canonical metrics in Kähler geometry*. Lectures in Mathematics ETH Zürich. Notes taken by Meike Akveld. Birkhäuser Verlag, Basel, pp. vi+101. MR: 1787650 (cit. on pp. 619, 633).

Received 2017-12-04.

SÉBASTIEN BOUCKSOM
CNRS-CMLS
ÉCOLE POLYTECHNIQUE
F-91128 PALAISEAU CEDEX
FRANCE
sebastien.boucksom@polytechnique.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (637–652)

AUTOMORPHISMS AND DYNAMICS: A LIST OF OPEN PROBLEMS

Serge Cantat

Abstract

We survey a few results concerning groups of regular or birational transformations of projective varieties, with an emphasis on open questions concerning these groups and their dynamical properties.

Contents

1 Introduction

1.1 Algebraic transformations. Let $\mathbf{k}$ be a field and d be a positive integer. Consider a smooth projective variety $X_{\mathbf{k}}$ of dimension d, defined over the field $\mathbf{k}$. Attached to $X_{\mathbf{k}}$ are two groups of algebraic transformations: its group of *birational transformations* $\mathrm{Bir}(X_{\mathbf{k}})$ and the subgroup of (biregular) *automorphisms* $\mathrm{Aut}(X_{\mathbf{k}})$. An element f of $\mathrm{Bir}(X_{\mathbf{k}})$ is defined by its graph $\mathfrak{G}_f \subset X_{\mathbf{k}} \times X_{\mathbf{k}}$; by definition, $\mathfrak{G}_f$ is an irreducible algebraic subvariety of dimension d such that the two projections $\mathfrak{G}_f \to X_{\mathbf{k}}$ have degree one, and in the

The author is grateful to the CNRS, Université Rennes 1, and foundation del Duca from the french Academy of Sciences, for their support, and expresses his warmest thanks to Eric Bedford, Jérémy Blanc, Yves de Cornulier, Julie Déserti, Jeffrey Diller, Igor Dolgachev, Romain Dujardin, Charles Favre, Vincent Guirardel, Stéphane Lamy, Federico Lo Bianco, Bjorn Poonen, Juan Souto, and Christian Urech for interesting discussions on the topics covered in this survey.
MSC2010: primary 14J50; secondary 37F99, 14E07, 14J28, 32M05, 37E30.
Keywords: Projective varieties, groups of automorphisms, transformation groups, complex dynamics, birational geometry.

case of automorphisms, the two projections are isomorphisms (in particular, they do not contract any algebraic subset of positive dimension onto a point).

For simplicity, unless otherwise specified, we assume that $\mathbf{k} = \mathbf{C}$ is the field of complex numbers. We denote by X the variety, with no reference to its field of definition, and by $X(\mathbf{C})$ its complex points; thus, $X(\mathbf{C})$ is also a compact, complex manifold of dimension d (and of real dimension $2d$). The group $\mathrm{Bir}(X)$ coincides with the group of bimeromorphic transformations of $X(\mathbf{C})$, and $\mathrm{Aut}(X)$ coincides with the subgroup of holomorphic diffeomorphisms.

A birational transformation $f : X \dashrightarrow X$ is a *pseudo-automorphism* if there exist two Zariski closed subsets Z and Z' of codimension ≥ 2 in X such that f induces an isomorphism from $X \setminus Z$ to $X \setminus Z'$. Equivalently, f and its inverse f^{-1} do not contract any hypersurface (they are “isomorphisms in codimension 1”). Pseudo-automorphisms constitute an important subgroup $\mathrm{Psaut}(X)$ of $\mathrm{Bir}(X)$ that contains $\mathrm{Aut}(X)$. As we shall see, automorphisms are quite rare in dimension ≥ 3; pseudo-automorphisms appear more frequently, for instance in the study of some special varieties, such as rational varieties and Calabi-Yau varieties.

1.2 Examples.

1.2.1 Projective spaces. Consider the projective space $\mathbb{P}^d_{\mathbf{k}}$ of dimension d. Its group of automorphisms is the group of linear projective transformations $\mathsf{PGL}_{d+1}(\mathbf{k})$. Its group of birational transformations is known as the Cremona group in d variables. In homogeneous coordinates $[x_1 : \ldots : x_{d+1}]$, every birational transformation f of $\mathbb{P}^d_{\mathbf{k}}$ can be written as

$$f[x_1 : \ldots : x_{d+1}] = [f_1 : \ldots : f_{d+1}] \tag{1}$$

where the f_i are homogeneous polynomials in the variables x_i, of the same degree, and without common factor of positive degree. In the affine coordinates $X_i = x_i/x_{d+1}$, it is defined by rational fractions

$$F_i(X_1, \ldots, X_d) = \frac{f_i(X_1, \ldots, X_d, 1)}{f_{d+1}(X_1, \ldots, X_d, 1)}. \tag{2}$$

When $d = 1$, $\mathrm{Bir}(\mathbb{P}^1_{\mathbf{k}})$ coincides with the group of automorphisms $\mathsf{PGL}_2(\mathbf{k})$. When $d \geq 2$, $\mathrm{Bir}(\mathbb{P}^d_{\mathbf{k}})$ is much bigger than $\mathsf{PGL}_{d+1}(\mathbf{k})$. In dimension 2, it contains all monomial transformations

$$(X_1, X_2) \mapsto (X_1^a X_2^b, X_1^c X_2^b) \tag{3}$$

with $ad - bc = \pm 1$, all transformations $(X_1, X_2) \mapsto (X_1, X_2 + h(X_1))$, for every $h \in \mathbf{k}(X_1)$, all linear projective transformations, hence all compositions of such maps. When

$\mathbf{k} = \mathbf{C}$, the Hénon map

$$(X_1, X_2) \mapsto (X_2, X_1 + X_2^2 + c) \tag{4}$$

provides a transformation of the plane with a rich dynamics (its topological entropy is equal to $\log(2)$ for every parameter c, it has infinitely many periodic points, ...).

1.2.2 Abelian varieties. Next, consider an elliptic curve $E = \mathbf{C}/\Lambda$, where Λ is a lattice in $\mathbf{C}$. The product $A = E^d$ is a complex torus of dimension d; it is also a projective variety, hence an example of an abelian variety. The group of birational transformations of A coincides with its group of automorphisms. It contains the group A itself, acting by translations, as well as the group $\mathsf{GL}_d(\mathbf{Z})$, acting by linear transformations on A (or more precisely on its universal cover $\mathbf{C}^d$, preserving the lattice Λ^d, hence also on A after taking the quotient by Λ^d).

1.2.3 Calabi-Yau varieties. As a third example, fix an integer $d \geq 1$, and consider a smooth hypersurface X in $(\mathbb{P}^1_{\mathbf{C}})^{d+1}$ which is defined in the open set $\mathbf{C}^{d+1}$ of $\mathbb{P}^1(\mathbf{C})^{d+1}$ by a polynomial equation $P(x_1, \ldots, x_{d+1}) = 0$ whose degree is equal to 2 with respect to each coordinate x_j. Geometrically, this means that for every index $1 \leq i \leq d+1$, the projection $\pi_i \colon X \to (\mathbb{P}^1_{\mathbf{C}})^d$ which forgets the i-th coordinate is a morphism of degree 2. The involution that permutes the two points in the fibers of π_i is a birational involution σ_i of X.

When $d \geq 2$, these hypersurfaces of degree $(2, 2, \ldots, 2)$ in $(\mathbb{P}^1_{\mathbf{C}})^{d+1}$ are examples of *Calabi-Yau varieties*: they are simply connected, they support a holomorphic d-form that does not vanish, and they do not split as the product of two varieties of smaller dimension. Thus, the involutions σ_i are pseudo-automorphisms because $\mathrm{Bir}(X) = \mathrm{Psaut}(X)$ for all Calabi-Yau varieties, as one sees by pulling back a non-vanishing holomorphic d-form.

In dimension ≤ 2, these pseudo-automorphisms are in fact regular automorphisms. In dimension 1, X is a curve of genus 1, and the involutions determine a dihedral group, acting by affine transformations of type $z \mapsto \pm z + a$ on the elliptic curve. In dimension $d = 2$, X is a K3 surface, and the involutions determine a large group of automorphisms of X: if the equation P is generic, then $\mathrm{Aut}(X) = \mathrm{Bir}(X)$ is generated by those three involutions, there are no relations between these involutions (they generate a group isomorphic to the free product $\mathbf{Z}/2\mathbf{Z} \star \mathbf{Z}/2\mathbf{Z} \star \mathbf{Z}/2\mathbf{Z}$), and the composition $\sigma_1 \circ \sigma_2 \circ \sigma_3$ has a rich dynamical behaviour (its topological entropy is $\log(9 + 4\sqrt{5})$).

For $d \geq 3$, the involutions σ_i are pseudo-automorphisms with indeterminacy points. For a generic choice of the equation P, the involutions σ_i generate $\mathrm{Psaut}(X)$, there are no relations between the σ_i, and $\mathrm{Aut}(X)$ is trivial (see Cantat and Oguiso [2015]).

1.3 Plan. Our goal is to review some important facts concerning the group $\mathrm{Aut}(X)$, the algebraic structure of its subgroups, and the dynamical properties of its elements. The emphasis is on open problems, in which algebraic geometry, group theory, and dynamics are simultaneously involved.

This article comprises two main parts. The first one concerns groups of automorphisms of smooth complex projective varieties, and their action on the cohomology of the variety. The second part concerns the dynamics: we focus on automorphisms with a dynamical behavior of low complexity because their study has been surprisingly neglected, while it offers interesting questions at the interface between dynamics and algebraic geometry.

We focus on $\mathrm{Aut}(X)$ for simplicity. As Section 1.2.3 shows, it would be better to work with pseudo-automorphisms, or even with birational transformations. In fact, most of the questions which are described below could be stated for pseudo-automorphisms of compact kähler manifolds (see Remark 2.1); and some of them concern birational transformations of projective varieties over an arbitrary field.

2 Groups of automorphisms

2.1 Automorphisms. Let X be a smooth complex projective variety. Its group of automorphisms is a complex Lie group (for the topology of uniform convergence), whose Lie algebra is the finite dimensional algebra of regular vector fields on X. But $\mathrm{Aut}(X)$ may have infinitely many components, as shown in Section 1.2.2 by the example of the abelian variety $(\mathbf{C}/\Lambda)^d$, where $\Lambda \subset \mathbf{C}$ is a lattice and $d \geq 2$.

Consider the action of $\mathrm{Aut}(X)$ on the cohomology $H^*(X;\mathbf{Z})$ of $X(\mathbf{C})$; this gives a linear representation

$$\mathrm{Aut}(X) \to \mathsf{GL}(H^*(X;\mathbf{Z})). \tag{5}$$

The connected component of the identity $\mathrm{Aut}(X)^0 \subset \mathrm{Aut}(X)$ acts trivially on the cohomology, and is therefore contained in the kernel of this representation. If an element $f \in \mathrm{Aut}(X)$ acts trivially on the second cohomology group $H^2(X;\mathbf{Z})$, it preserves the cohomology class of a kähler form κ on X (resp. the first Chern class of an ample line bundle L on X); this means that the volume (resp. degree) of the graph of f is uniformly bounded on the kernel of the representation (5). Since subvarieties with a fixed volume (resp. degree) form a bounded family, one can use Douady spaces (or Hilbert Schemes) to obtain the following fact: *the connected component of the identity in* $\mathrm{Aut}(X)^0 \subset \mathrm{Aut}(X)$ *has finite index in the kernel of the linear representation* $\mathrm{Aut}(X) \to \mathsf{GL}(H^2(X;\mathbf{Z}))$ (see Grothendieck [1995] and Lieberman [1978] for a proof).

Thus, the group $\mathrm{Aut}(X)$ splits into two basic parts: its neutral component $\mathrm{Aut}(X)^0$, and its discrete image

$$\mathrm{Aut}(X)^* \subset \mathsf{GL}(H^*(X;\mathbf{Z})). \tag{6}$$

The group of connected components $\mathrm{Aut}(X)/\mathrm{Aut}(X)^0$ is an extension of $\mathrm{Aut}(X)^*$ by a finite group.

Remark 2.1. One may replace $H^2(X;\mathbf{Z})$ by the subgroup $\mathrm{N}^1(X)$ generated by cohomology classes of algebraic hypersurfaces of X (or more precisely by the Poincaré dual of their homology classes). The kernel of the representation $\mathrm{Aut}(X) \mapsto \mathsf{GL}(\mathrm{N}^1(X))$ is, again, equal to a finite extension of $\mathrm{Aut}(X)^0$. Doing so, it is possible to phrase some of the following questions for varieties which are defined over fields of arbitrary characteristic.

One may also replace $\mathrm{Aut}(X)$ by $\mathrm{Psaut}(X)$. Indeed, since pseudo-automorphisms are isomorphisms in codimension 1, the group $\mathrm{Psaut}(X)$ acts linearly on $\mathrm{N}^1(X)$ (and $H^2(X;\mathbf{Z})$ when $\mathbf{k} = \mathbf{C}$). Thus, all questions concerning the action of $\mathrm{Aut}(X)$ on $H^2(X;\mathbf{Z})$ can be stated for the action of $\mathrm{Psaut}(X)$ on $\mathrm{N}^1(X)$ (even when the characteristic of $\mathbf{k}$ is positive).

2.2 The realization problem. Two main problems arise: given a connected algebraic group G, does there exist a projective variety X such that $\mathrm{Aut}(X)^0$ is isomorphic to G as an algebraic group ? Given a subgroup Γ of $\mathsf{GL}_n(\mathbf{Z})$, for some $n \geq 2$, does there exist a projective variety X and an isomorphism of groups $\Gamma \simeq \mathrm{Aut}(X)^*$? There is also a third, less interesting problem, which asks which pairs (G, Γ) may be simultaneously realized as the connected component and the discrete part of $\mathrm{Aut}(X)$, for some variety X.

The first problem has been solved by Brion in the following strong sense: *any connected algebraic group G over a perfect field is the neutral component of the automorphism group scheme of some normal projective variety X; if the characteristic of the field is 0, one can moreover assume that X is smooth of dimension* $\dim(X) = 2\dim(G)$ (see Brion [2014]; see also Winkelmann [2004] for Kobayashi hyperbolic manifolds).

In the following paragraphs, we focus on the discrete, countable group $\mathrm{Aut}(X)^*$.

2.3 Linear groups and their Zariski closure. Not much has been proven yet concerning the description of the groups $\mathrm{Aut}(X)^*$. To simplify the exposition, we shall say that a group Γ is *abstractly realizable* as a group of automorphisms (in dimension d) if there exist a projective variety X (of dimension d) such that $\mathrm{Aut}(X)$ is isomorphic to Γ (as abstract groups).

Remark 2.2. The group $\mathsf{GL}_n(\mathbf{Z})$ acts on the abelian variety $(\mathbf{C}/\Lambda)^n$ for every lattice $\Lambda \subset \mathbf{C}$. Blowing-up the origin, one gets a new variety X with $\mathrm{Aut}(X)^0 = \{\mathrm{id}_X\}$ and with

$\mathsf{GL}_n(\mathbf{Z}) \simeq \mathrm{Aut}(X)$ if Λ is generic. Thus, every subgroup Γ of $\mathsf{GL}_n(\mathbf{Z})$ acts faithfully on a projective variety of dimension n whose group of automorphisms is discrete. This does not say that Γ is abstractly realizable.

Remark 2.3. A necessary condition for a countable group Γ to be realizable is that Γ admits a linear integral representation $\Gamma \to \mathsf{GL}_m(\mathbf{Z})$ with finite kernel. This is not a sufficient condition: *when $n \geq 2$, there is a finite, cyclic, and central extension of the symplectic group $\mathsf{Sp}_{2n}(\mathbf{Z})$ that does not act faithfully by birational transformations on any complex projective variety* (see Cantat and Xie [2015]).

Instead of looking at the abstract notion of realizability, one may also add some rigidity in the definition; this may be done as follows. Say that a group $\Gamma \subset \mathsf{GL}_m(\mathbf{Z})$ is *strongly realizable* as a group of automorphisms if there is a smooth complex projective variety X, and a linear algebraic representation

$$\varphi\colon \mathsf{GL}_m(\mathbf{R}) \to \mathsf{GL}(H^2(X;\mathbf{R})) \tag{7}$$

which is defined over $\mathbf{Z}$, such that $\mathrm{Aut}(X)^*$ coincides with $\varphi(\Gamma)$. As in Remark 2.1, one may replace $H^2(X;\mathbf{Z})$ by $\mathrm{N}^1(X)$ to define a notion of strong realizability by pseudo-automorphisms, defined over any field $\mathbf{k}$.

Remark 2.4. The cohomology group $H^2(X;\mathbf{C})$ splits into the direct sum of the Dolbeault groups $H^{p,q}(X;\mathbf{C})$ with $p+q=2$. This splitting, the intersection form and the cone of all Kähler classes are $\mathrm{Aut}(X)^*$-invariant (see Griffiths and Harris [1978]). Hence, it would be too much to require φ to be an isomorphism.

There are only countably many distinct groups $\mathrm{Aut}(X)^*$. This follows from the fact that, up to isomorphism, there are only countably many pairwise non-isomorphic extensions of $\mathbf{Q}$ which are finitely generated. On the opposite, for every integer $n \geq 4$, there are uncountably many, pairwise non-isomorphic, subgroups of $\mathsf{GL}_n(\mathbf{Z})$. Thus a counting argument shows that most subgroups Γ of $\mathsf{GL}_n(\mathbf{Z})$ are not realizable as groups of automorphisms. On the other hand, there are only countably many subgroups of $\mathsf{GL}_n(\mathbf{Z})$ which are finitely generated.

Question 2.5. Which subgroups of $\mathsf{GL}_n(\mathbf{Z})$ are abstractly (resp. strongly) realizable as groups of automorphisms ? Is every finitely generated subgroup $\Gamma \subset \mathsf{GL}_n(\mathbf{Z})$ abstractly (resp. strongly) realizable as a group of automorphisms ?

Recently, Lesieutre proved that *if $\mathbf{k}$ is a field of characteristic 0, or a field which is not algebraic over its prime field, then there is a smooth, 6-dimensional projective variety X over $\mathbf{k}$ such that* $\mathrm{Aut}(X)^0$ *is trivial and* $\mathrm{Aut}(X)^*$ *is not finitely generated* (see Lesieutre [2017] and Dinh and Oguiso [2017] for another example). This shows that it is somewhat artificial to assume that Γ is finitely generated.

Remark 2.6. The following problem remains open: does there exist a rational surface S such that $\mathrm{Aut}(S)$ is discrete but not finitely generated ? More precisely, can one find such an example which is a minimal model for the pair $(S, \mathrm{Aut}(S))$, meaning that if $\pi : S \to S'$ is a birational morphism and $\pi \circ \mathrm{Aut}(S) \circ \pi^{-1} \subset \mathrm{Aut}(S')$ then π is an isomorphism ?

In Question 2.5, the dimension of X is not specified; indeed, $\dim(X)$ must a priori be large with respect to n. Changing viewpoint, one may fix the dimension, or fix the type of variety one considers, and try to find constraints on the subgroups $\mathrm{Aut}(X)^* \subset \mathsf{GL}(H^2(X;\mathbf{Z}))$.

Question 2.7. Fix a dimension $d \geq 2$. Define $G^0(X)$ as the neutral component of the Zariski closure of $\mathrm{Aut}(X)^*$ in $\mathsf{GL}(H^2(X;\mathbf{R}))$. What kind of linear algebraic groups $G^0(X)$ do we obtain in this way, when X runs over all possible Calabi-Yau varieties of dimension d ?

Again, by Remark 2.1, the group $\mathrm{Aut}(X)$ may be replaced by the group $\mathrm{Psaut}(X) = \mathrm{Bir}(X)$ and $H^2(X;\mathbf{R})$ by $\mathrm{N}^1(X) \otimes_{\mathbf{Z}} \mathbf{R}$ in this question. For a nice example, see Cantat and Oguiso [2015]. For complex projective K3 surfaces, the groups $G^0(X)$ that one gets are connected components of $\mathsf{SO}_{1,m}(\mathbf{R})$ for some $m \leq 19$, or abelian groups $\mathbf{R}^m$ of rank $m \leq 18$. This comes from Hodge index theorem.

Question 2.5 is stated for Calabi-Yau manifolds because they form one of the most interesting classes of examples, but it may be stated for other classes, for instance for rational varieties. And instead of looking at the action of $\mathrm{Aut}(X)$ on $H^2(X;\mathbf{R})$, one may also consider its action on every Dolbeault subgroup $H^{p,q}(X;\mathbf{C})$ (see Cantat and Dolgachev [2012], Cantat and Zeghib [2012], and Zhang [2013]).

2.4 Real algebraic variation. Consider now a smooth real projective variety $X_{\mathbf{R}}$ of dimension d. Assume that $X(\mathbf{R})$ is non-empty, and fix one of the connected components $S \subset X(\mathbf{R})$ (for the euclidean topology); then S is a closed connected manifold of dimension d. The automorphisms of X which are defined over $\mathbf{R}$ form a subgroup $\mathrm{Aut}(X_{\mathbf{R}})$ of $\mathrm{Aut}(X_{\mathbf{C}})$. A finite index subgroup $\mathrm{Aut}(X_{\mathbf{R}}; S)$ of $\mathrm{Aut}(X_{\mathbf{R}})$ fixes S, and we get a restriction morphism

$$\mathrm{Aut}(X_{\mathbf{R}}; S) \to \mathrm{Diff}^{\infty}(S) \tag{8}$$

where $\mathrm{Diff}^{\infty}(S)$ is the group of $\mathcal{C}^{\infty}$-diffeomorphisms of S. Denoting by $\mathrm{Mod}(S)$ the modular group of S, i.e. the group of connected components of $\mathrm{Diff}^{\infty}(S)$ (see Farb and Margalit [2012]), one gets a homomorphism

$$\alpha_S : \mathrm{Aut}(X_{\mathbf{R}}; S) \to \mathrm{Mod}(S). \tag{9}$$

What can be said on the image of this homomorphism ?

The best is to start with surfaces. First, there are explicit examples of automorphisms f of rational and K3 surfaces for which the mapping class $\alpha_S(f)$ is interesting (see Bedford and Kim [2009] and Moncet [2008]). Second, Kollár and Mangolte obtained the following beautiful result. Fix any smooth real surface $X_{\mathbf{R}}$, which is rational (over $\mathbf{R}$). Then, consider the subgroup $\mathrm{Bir}^\infty(X_{\mathbf{R}})$ of $\mathrm{Bir}(X_{\mathbf{R}})$ defined by the following property: f is in $\mathrm{Bir}^\infty(X_{\mathbf{R}})$ if and only if f and its inverse f^{-1} have no real indeterminacy point – all its indeterminacy points come in complex conjugate pairs. By restriction to $X(\mathbf{R})$, each element f of $\mathrm{Bir}^\infty(X_{\mathbf{R}})$ provides a diffeomorphism of $X(\mathbf{R})$, and therefore a mapping class $\alpha_{X(\mathbf{R})}(f) \in \mathrm{Mod}(X(\mathbf{R}))$. In Kollár and Mangolte [2009] it is proved that *this homomorphism* $\mathrm{Bir}^\infty(X_{\mathbf{R}}) \to \mathrm{Mod}(X(\mathbf{R}))$ *is surjective*.

Question 2.8. Does there exist a real projective K3 surface X such that (i) $X(\mathbf{R})$ is connected and of genus $g \geq 2$ and (ii) the image of $\alpha_{X(\mathbf{R})}\colon \mathrm{Aut}(X_{\mathbf{R}}) \to \mathrm{Mod}(X(\mathbf{R}))$ is surjective ?

I suspect that the answer to this question is negative. Indeed, there may exist a mapping class ψ of the closed, orientable surface of genus 2 with the following property: given any automorphism f of a real K3 surface $X_{\mathbf{R}}$ with an f-invariant connected component $S \subset X(\mathbf{R})$ of genus 2, $\alpha_S(f)$ is not conjugate to ψ.

2.5 Dynamical degrees. Consider an automorphism f of a smooth complex projective variety X. The characteristic polynomial $\chi_f(t)$ of

$$f^*\colon H^2(X;\mathbf{Z}) \to H^2(X;\mathbf{Z}) \tag{10}$$

is an element of $\mathbf{Z}[t]$; it is monic, and its constant term is ± 1. Denote by $\lambda_1(f)$ the largest absolute value of a root of $\chi_f(t)$. The invariance of the Hodge decomposition and of the Kähler cone implies that $\lambda_1(f)$ is in fact one of the roots of $\chi_f(t)$ and, as such, is an algebraic integer. This number $\lambda_1(f)$ is called the first *dynamical degree* of f (subsequent dynamical degrees are obtained by looking at the action of f on the groups $H^{p,p}(X;\mathbf{R})$ with $p > 1$). Now, fix the dimension d of the variety, and define the following set of real integers:

$$\mathfrak{S}_1(d) = \{\lambda_1(f);\ f \in \mathrm{Aut}(X) \text{ for } X \text{ smooth, projective, of dimension } d\}. \tag{11}$$

In dimension 1, $\mathfrak{S}_1(1) = \{1\}$, but in dimension $d \geq 2$, $\mathfrak{S}_1(d)$ is an infinite countable set of algebraic integers. In dimension 2, the Hodge index theorem shows that $\mathfrak{S}_1(2) \setminus \{1\}$ contains only reciprocal quadratic integers and Salem numbers; and its first derive set is non empty: for example the golden mean is an increasing limit of elements of $\mathfrak{S}_1(2)$ (see Diller and Favre [2001]). A deeper result says that *every strictly decreasing sequence of elements of* $\mathfrak{S}_1(2)$ *is finite*; in particular, *given any* $\alpha \in \mathfrak{S}_1(2)$, *there is a real number*

$\epsilon(\alpha) > 0$ *such that* $]\alpha, \alpha+\epsilon(\alpha)[$ *does not intersect* $\mathfrak{S}_1(2)$ (see McMullen [2002a, 2007] and Blanc and Cantat [2016]). In particular, many Salem numbers are not contained in $\mathfrak{S}_1(2)$. The only general constraint known on the elements of $\mathfrak{S}_1(3)$ is due to Lo Bianco: *if* $\alpha \in \mathfrak{S}_1(3)$*, there are at most* 6 *distinct moduli* $|\alpha_i|$ *for the Galois conjugates* α_i *of* α (whatever the degree of the algebraic integer α, see Lo Bianco [2014]).

Question 2.9. Is $\mathfrak{S}_1(3)$ dense in $[1, +\infty[$? Or, on the contrary, does there exist $\epsilon > 0$ such that $\mathfrak{S}_1(3) \cap]1, 1+\epsilon]$ is empty ?

In this question, on may replace $\mathfrak{S}_1(3)$ by the set of dynamical degrees of all *birational* transformations of projective threefolds; here, by definition, the first dynamical degree of a birational transformation $f \in \mathrm{Bir}(X)$ is defined by the limit

(12) $$\lambda_1(f) = \lim_{n\to+\infty} \left((f^n)^*(H)\cdot H^{d-1}\right)^{1/n}$$

where H is some hyperplane section of X, d is the dimension of X, and the integer $(f^n)^*(H)\cdot H^{d-1}$ is the intersection product of the total transform $(f^n)^*(H)$ with $d-1$ copies of H. This limit exists and does not depend on H (see Dinh and Sibony [2005], Dang [2017], and Truong [2016], and Section 3.3 below).

3 Dynamics with low complexity

3.1 Entropy. Given a continuous transformation g of a compact metric space M, the topological *entropy* $\mathrm{h}_{top}(g; M)$ is a measure of the complexity of the dynamics of $g: M \to M$. To define it, fix a scale $\epsilon > 0$ at which you observe the dynamics. Given a period of observation $N \in \mathbf{Z}_+$, count the maximum number of orbits $(x, g(x), \ldots, g^{N-1}(x))$ that are pairwise distinct at scale ϵ, the orbit of x being distinguishable from the orbit of y if the distance from $g^k(x)$ to $g^k(y)$ is larger than ϵ for at least one time $0 \le k < N$. This number of orbits $\mathrm{Orb}(N;\epsilon)$ typically grows exponentially fast with N; thus, one defines $h(g;\epsilon)$ as the supremum limit of $\frac{1}{N}\log(\mathrm{Orb}(N;\epsilon))$ and the entropy $\mathrm{h}_{top}(g; M)$ of g as the limit of $h(g;\epsilon)$ as ϵ goes to 0, i.e. as our observations become arbitrarily accurate.

Topological entropies are hard to estimate for continuous or smooth maps of manifolds. But *the topological entropy of an endomorphism* f *of a smooth complex projective variety* X *is equal to the logarithm of the spectral radius of* $f^*: H^*(X;\mathbf{C}) \to H^*(X;\mathbf{C})$ (the action of f on the cohomology). This wonderful result is due to Gromov and Yomdin (see Gromov [2003, 1987] and Yomdin [1987], and Dinh and Sibony [2005] for an upper bound when f is a rational map). For instance, the entropy is invariant under deformation: if F is an automorphism of a variety $\mathfrak{X}$ that preserves a fibration $\pi: \mathfrak{X} \to B$ with smooth projective fibers and the fibration is locally trivial topologically, then the automorphisms induced by F on the fibers of π have the same entropy.

Endomorphisms with positive entropy have been studied in detail, with special focus on endomorphisms of the projective space $\mathbb{P}^k_{\mathbf{C}}$ and automorphisms of surfaces. We refer to Cantat [2014] and Dinh and Sibony [2010a] for survey papers on the subject. Here, we consider the opposite edge of the spectrum: instead of looking at automorphisms with chaotic dynamics, we ask for a description of automorphisms with dynamics of low complexity.

3.2 Invariant fibrations. Now consider an automorphism f of a smooth complex projective variety X with entropy equal to 0. Then, by the Gromov-Yomdin theorem, the eigenvalues of $f^* \in \mathsf{GL}(H^*(X;\mathbf{Z}))$ all have modulus 1, and being algebraic integers, they must be roots of unity. Changing f in a positive iterate, we may assume that f^* is *unipotent*: all its eigenvalues are equal to 1.

If f^* is the identity, then some further iterate is contained in $\mathrm{Aut}(X)^0$ (see Section 2.1). The dynamics of such an automorphism is well understood. Thus, one may assume that f^* is non-trivial and unipotent; equivalently, the sequence $\| (f^n)^* \|$ grows polynomially quickly with n, as n^k for some $k \geq 1$.

When $\dim(X) = 2$, Gizatullin proved the following: *if f^* is unipotent and $\neq \mathrm{Id}$, then f preserves a (singular) fibration by curves of genus 1 and the growth of $\| (f^n)^* \|$ is quadratic* (see Gizatullin [1980], Diller and Favre [2001], and Cantat [2001]). For instance, if $\pi : X \to B$ is a genus 1 fibration of the surface X with two sections, then the translations along the fibers that map the first section to the second one determine an algebraic transformation of X: this is often an automorphism such that $\| (f^n)^* \|$ grows quadratically. The following question asks whether Gizatullin's classification can be extended to higher dimension.

Question 3.1. Let X be a smooth complex projective variety of dimension 3. Let f be an automorphism of X, such that the linear transformation $f^* \in \mathsf{GL}(H^*(X;\mathbf{C}))$ is non-trivial and unipotent. Does f permute the fibers of a non-trivial meromorphic fibration $\pi : X \dashrightarrow B$?

By "non-trivial meromorphic fibration", I mean that $0 < \dim(B) < \dim(X)$ and π is a dominant rational map; and f permutes the fibers of π if there is a birational transformation f_B of B such that $\pi \circ f = f_B \circ \pi$.

When f is an automorphism of a smooth complex projective variety of dimension 3 and f^* is a non-trivial unipotent matrix, then $\| (f^*)^n \|$ *grows like n^k with $k \in \{2, 4\}$* (see Lo Bianco [2014]). But beside this general statement, not much is known.

One of the first examples to look at is the case of Calabi-Yau varieties. So, let f be an automorphism of a Calabi-Yau variety X of dimension 3, with a non-trivial unipotent action on the cohomology. Consider a kähler form κ on X, with cohomology class $[\kappa] \in H^2(X;\mathbf{R})$; then $n^{-k}(f^n)^*[\kappa]$ converges towards an f-invariant nef class. Is this class related to an f-invariant fibration on X ? This is a version of the abundance conjecture (see Lazić, Oguiso, and Peternell [2016]), precisely in a case which is not solved yet, but with the additional presence of an automorphism.

3.3 Degree growth. If an automorphism satisfies $\lambda_1(f) = 1$, then an iterate of f^* is unipotent, and $\| (f^n)^* \|$ grows polynomially with n. Now, consider a birational transformation $f : X \dashrightarrow X$ of a complex projective variety X. Fix a polarization H of X and define the degree of f with respect to H by the intersection product

$$\deg_H(f) = f^*(H) \cdot H^{d-1} \tag{13}$$

where d is the dimension of X and $f^*(H)$ is the total transform of H by f^{-1}. Changing H into another polarization H', the notion of degree is only perturbed by a bounded multiplicative error: there exists a positive constant c such that

$$\frac{1}{c}\deg_{H'}(f) \leq \deg_H(f) \leq c \deg_{H'}(f) \tag{14}$$

for all $f \in \mathrm{Bir}(X)$ (see Dinh and Sibony [2005], Dang [2017], and Truong [2016]). When $X = \mathbb{P}^d_{\mathbf{C}}$ and $H \subset \mathbb{P}^d_{\mathbf{C}}$ is a hyperplane, then $\deg_H(f)$ is just the degree of the polynomials that define f in homogeneous coordinates (see Section 1.2.1). Instead of assuming that the entropy of f is zero, one may now assume that $\lambda_1(f) = 1$, i.e. that $\deg_H(f^n)$ does not grow exponentially fast with n.

Question 3.2. Let X be a smooth projective variety of dimension d, together with a polarization H. Let f be a birational transformation of X with $\lambda_1(f) = 1$. Does the sequence $\deg_H(f^n)$ grow polynomially ? If not bounded, does this sequence grow at least linearly ?

In other words, can one construct a birational transformation f, say, of $\mathbb{P}^3_{\mathbf{C}}$, such that the sequence $\deg_H(f^n)$ grows like $\exp(\sqrt{n})$, or like $n^2\log(n)$, or like $n^{1/3}$? This is already an open problem for polynomial automorphisms of the affine space $\mathbb{A}^3_{\mathbf{C}}$. One may also ask, as in Question 3.1, whether the equality $\lambda_1(f) = 1$ implies the existence of a non-trivial invariant fibration. While these questions are fully understood in dimension 2 (see Cantat [2013, 2015], Diller and Favre [2001], and Gizatullin [1980]) almost nothing is known in higher dimension (see Urech [2017], as well as Déserti [2017] and Lin [2013] for interesting examples).

This type of question is, in fact, directly related to the algebraic structure of the group $\mathrm{Bir}(X)$. For instance, one says that an element f of $\mathrm{Bir}(X)$ is *distorted* if there exists a finite set $S \subset \mathrm{Bir}(X)$ such that the n-th iterate f^n is a product $g_1 \circ \cdots \circ g_{\ell(n)}$ of elements $g_i \in S$ with $\liminf(\ell(n)/n) = 0$ (see Calegari and Freedman [2006]). Distorted elements satisfy $\lambda_1(f) = 1$. In dimension 2, an element f of $\mathrm{Bir}(\mathbb{P}^2_{\mathbf{C}})$ is distorted if and only if some positive iterate f^N is conjugate to an element of $\mathrm{Aut}(\mathbb{P}^2_{\mathbf{C}})$. Any progress on Questions 3.1 and 3.2 would certainly help classify distorted elements in higher dimension.

3.4 Minimal transformations. A continuous transformation g of a compact metric space M is *minimal* if each of its orbits are dense; equivalently, the only g-invariant closed subsets are M and the empty set. For an algebraic example, consider a complex torus $A = \mathbf{C}^k/\Lambda$, with Λ a cocompact lattice in $\mathbf{C}^k$; then, every totally irrational translation on A is minimal. For another example, take $k = 2$ and $\Lambda = \Lambda_0 \times \Lambda_0$, with Λ_0 a lattice in $\mathbf{C}$. Choose a point a in $\mathbf{C}/\Lambda_0$ such that each orbit of $x \mapsto x + a$ is dense in $\mathbf{C}/\Lambda_0$, and consider the following automorphism of $\mathbf{C}^2/\Lambda$:

$$f(x, y) = (x + a, y + x) \qquad \mathrm{mod}\ (\Lambda). \tag{15}$$

Furstenberg shows in Furstenberg [1961] that every orbit of f is dense in the abelian surface $\mathbf{C}^2/\Lambda$. In fact, Furstenberg proves more: *f preserves the Haar measure on the torus and this is the unique f-invariant probability measure*. Thus, there are minimal automorphisms with no positive iterate in $\mathrm{Aut}(X)^0$ (in Furstenberg's example, the action of f on the cohomology is non-trivial and unipotent).

A minimal automorphism of a curve is a translation on an elliptic curve. In dimension 2, one also proves easily that *any complex projective surface X with a minimal automorphism is an abelian surface*. It would be great to get a similar result in higher dimension:

Question 3.3. Let X be a complex projective variety of dimension 3. Suppose that there is a minimal automorphism f on X. Is X an abelian variety ?

If f is minimal it does not have any periodic orbit. With the Lefschetz formula and Gromov-Yomdin theorem, one sees that the entropies of minimal automorphisms of complex projective varieties vanish. Now, if X has dimension 3, the holomorphic Lefschetz formula says that $h^{1,0}(X) > 0$ or $h^{3,0}(X) > 0$. If $h^{1,0}(X) > 0$, one may use the Albanese morphism to reduce the complexity of the dynamics of f (see Cantat [2010]). If $h^{3,0}(X) > 0$, there is an f-invariant holomorphic 3-form on X; since f does not preserve any strict Zariski closed set (by minimality), this form does not vanish. Thus, the most interesting case is when X is a Calabi-Yau variety of dimension 3 with Euler characteristic equal to 0, and the action of f^m on the cohomology is unipotent for some $m > 0$ (as in Section 3.2).

3.5 Fatou components and real algebraic varieties.

3.5.1 Fatou components on K3 surfaces. Consider an automorphism f of a K3 surface X. Assume that the topological entropy of f on the complex surface $X(\mathbf{C})$ is positive: $h_{top}(f;\mathbf{C}) = \log(\lambda_1(f)) > 0$ where $\lambda_1(f)$ is, by the Gromov-Yomdin theorem, the largest eigenvalue of f on the cohomology of $X(\mathbf{C})$. In that case, f has infinitely many periodic points, and those points equidistribute toward an f-invariant probability measure μ_f (see Cantat [2001, 2014] and Dujardin [2006]). More precisely, denote by $\mathrm{Per}(f;N)$ the set of isolated periodic points of f or period N. Then, the sequence of probability measures

$$\frac{1}{|\mathrm{Per}(f;N)|}\sum_{z\in \mathrm{Per}(f;N)} \delta_z \tag{16}$$

converges toward an f-invariant probability measure μ_f with interesting dynamical properties (this is the unique measure of maximal entropy of f). The same property holds if one replaces $\mathrm{Per}(f;N)$ by the set of saddle periodic points of period N, i.e. by points $x \in X(\mathbf{C})$ such that $f^N(x) = x$ and the differential Df^N_x has two eigenvalues α and β with

$$|\alpha| > 1 > |\beta|; \tag{17}$$

indeed, the number of isolated periodic points and of saddle periodic points grow at the same speed, namely like $\lambda_1(f)^N$.

The measure μ_f is, in general, singular with respect to the Lebesgue measure on $X(\mathbf{C})$. *If μ_f is absolutely continuous with respect to the Lebesgue measure, then (X, f) is a Kummer example: this means that X is a quotient of an abelian surface A and f is induced by a linear automorphism of A* (see Cantat and Dupont [2015]).

But not much is known about the support of the measure μ_f. There may a priori exist a region $\mathcal{U} \subset X(\mathbf{C})$ which is f-invariant and on which the dynamics of f is far from being chaotic. More precisely, define the *Fatou set* $\mathrm{Fat}(f)$ as follows: $x \in X(\mathbf{C})$ is in the Fatou set if there is an open neighborhood $\mathcal{V}$ of x such that the sequence of iterates $(f^m)_{m\in\mathbf{Z}}$ forms a normal family on $\mathcal{V}$, in the sense of Montel. This determines an f-invariant open subset $\mathrm{Fat}(f) \subset X(\mathbf{C})$ on which μ_f vanishes. It is not known yet whether the Fatou set is always empty for automorphisms of complex projective K3 surfaces with positive entropy. There are examples of non-empty Fatou sets on non-projective K3 surfaces and rational surfaces (see McMullen [2002b] and Bedford and Kim [2009]).

Question 3.4. Does there exist a complex projective K3 surface X with an automorphism $f : X \to X$ such that the topological entropy of f is positive and the Fatou set of f is not empty ?

While the theory of closed positive currents, the Hodge theory, and the Pesin theory of smooth dynamical systems may be combined to study the chaotic part of the dynamics from a stochastic viewpoint (see Bedford, Lyubich, and Smillie [1993], Cantat [2001, 2014], De Thélin and Dinh [2012], and Dinh and Sibony [2010b] for instance), not much is known concerning the topological properties of the dynamics, such as the existence of non-empty Fatou components or dense orbits. Thus, Question 3.4 is at the borderline of our knowledge.

3.5.2 Dynamics on real K3 surfaces. Now, add to the hypotheses that X and f are defined over the field of real numbers $\mathbf{R}$, and $X(\mathbf{R})$ is not empty. Fix a connected component S of $X(\mathbf{R})$, and replace f by some positive iterate so that f now preserves S. We obtain two dynamical systems: the complex dynamics given by $f : X(\mathbf{C}) \to X(\mathbf{C})$, and the real dynamics given by $f : S \to S$. The topological entropy of f on S is bounded above by the topological entropy of f on $X(\mathbf{C})$. For rational surfaces, it may happen that these two numbers are equal (and positive, see Bedford and Kim [2009]); in that case most periodic points of f are indeed contained in $X(\mathbf{R})$. For K3 surfaces, there is no example yet with

$$\mathrm{h}_{top}(f; X(\mathbf{R})) > \frac{1}{2}\mathrm{h}_{top}(f; X(\mathbf{C})). \tag{18}$$

At the opposite edge of the possibilities, there is no known example for which $\mathrm{h}_{top}(f; X(\mathbf{C}))$ is positive while $\mathrm{h}_{top}(f; S) = 0$ for some connected component of $X(\mathbf{R})$ (see Moncet [2012]).

Question 3.5. Does there exist a real projective K3 surface $X_{\mathbf{R}}$ with an automorphism $f : X_{\mathbf{R}} \to X_{\mathbf{R}}$ such that $\lambda_1(f) > 1$, $X(\mathbf{R})$ is not empty, and one of the following property occurs

(1) the entropy of f on $X(\mathbf{R})$ is equal to the entropy of f on $X(\mathbf{C})$?

(2) the entropy of f on some component S of $X(\mathbf{R})$ vanishes ?

(3) a connected component S of $X(\mathbf{R})$ is contained in the Fatou set of f ?

Of course, a positive answer to the third of these questions would imply a positive answer to the second question and to Question 3.4. To obtain a positive answer to the first, it would be sufficient to find a component $S \subset X(\mathbf{R})$ which is f invariant and such that the mapping class $\alpha_S(f)$ has a stretching factor equal to $\lambda_1(f)$ on the fundamental group $\pi_1(S)$ (see Farb and Margalit [2012]); thus, Question 3.5 is also related to Section 2.4.

References

Eric Bedford and Kyounghee Kim (2009). "Dynamics of rational surface automorphisms: linear fractional recurrences". *J. Geom. Anal.* 19.3, pp. 553–583. MR: 2496566 (cit. on pp. 644, 649, 650).

Eric Bedford, Mikhail Lyubich, and John Smillie (1993). "Polynomial diffeomorphisms of $\mathbf{C}^2$. IV. The measure of maximal entropy and laminar currents". *Invent. Math.* 112.1, pp. 77–125. MR: 1207478 (cit. on p. 650).

Jérémy Blanc and Serge Cantat (2016). "Dynamical degrees of birational transformations of projective surfaces". *J. Amer. Math. Soc.* 29.2, pp. 415–471. MR: 3454379 (cit. on p. 645).

Michel Brion (2014). "On automorphisms and endomorphisms of projective varieties". In: *Automorphisms in birational and affine geometry*. Vol. 79. Springer Proc. Math. Stat. Springer, Cham, pp. 59–81. MR: 3229345 (cit. on p. 641).

Danny Calegari and Michael H. Freedman (2006). "Distortion in transformation groups". *Geom. Topol.* 10. With an appendix by Yves de Cornulier, pp. 267–293. MR: 2207794 (cit. on p. 647).

Serge Cantat (2001). "Dynamique des automorphismes des surfaces $K3$". *Acta Math.* 187.1, pp. 1–57. MR: 1864630 (cit. on pp. 646, 649, 650).

– (2010). "Quelques aspects des systèmes dynamiques polynomiaux: existence, exemples, rigidité". In: *Quelques aspects des systèmes dynamiques polynomiaux*. Vol. 30. Panor. Synthèses. Soc. Math. France, Paris, pp. 13–95. MR: 2932433 (cit. on p. 648).

– (2013). "The Cremona group in two variables". In: *European Congress of Mathematics*. Eur. Math. Soc., Zürich, pp. 211–225. MR: 3469123 (cit. on p. 647).

– (2014). "Dynamics of automorphisms of compact complex surfaces". In: *Frontiers in complex dynamics*. Vol. 51. Princeton Math. Ser. Princeton Univ. Press, Princeton, NJ, pp. 463–514. MR: 3289919 (cit. on pp. 646, 649, 650).

– (2015). "The Cremona groups". *Proceedings of the 2015 Summer Institute in Algebraic Geometry* to appear, pp. 1–48 (cit. on p. 647).

Serge Cantat and Igor Dolgachev (2012). "Rational surfaces with a large group of automorphisms". *J. Amer. Math. Soc.* 25.3, pp. 863–905 (cit. on p. 643).

Serge Cantat and Christophe Dupont (2015). "Automorphisms of surfaces: Kummer rigidity and measure of maximal entropy". *preprint*, pp. 1–68 (cit. on p. 649).

Serge Cantat and Keiji Oguiso (2015). "Birational automorphism groups and the movable cone theorem for Calabi-Yau manifolds of Wehler type via universal Coxeter groups". *Amer. J. Math.* 137.4, pp. 1013–1044. MR: 3372313 (cit. on pp. 639, 643).

Serge Cantat and Junyi Xie (2015). "Algebraic actions of discrete groups: The p-adic method". *preprint*, pp. 1–52 (cit. on p. 642).

Serge Cantat and Abdelghani Zeghib (2012). "Holomorphic actions, Kummer examples, and Zimmer program". *Ann. Sci. Éc. Norm. Supér. (4)* 45.3, pp. 447–489. MR: 3014483 (cit. on p. 643).

Nguyen-Bac Dang (2017). "Degrees of Iterates of Rational Maps on Normal Projective Varieties". arXiv: 1701.07760 (cit. on pp. 645, 647).

Henry De Thélin and Tien-Cuong Dinh (2012). "Dynamics of automorphisms on compact Kähler manifolds". *Adv. Math.* 229.5, pp. 2640–2655. MR: 2889139 (cit. on p. 650).

Julie Déserti (2017). "Degree growth of polynomial automorphisms and birational maps: some examples". *Eur. J. Math., to appear*, pp. 1–10 (cit. on p. 647).

Jeffrey Diller and Charles Favre (2001). "Dynamics of bimeromorphic maps of surfaces". *Amer. J. Math.* 123.6, pp. 1135–1169. MR: MR1867314 (cit. on pp. 644, 646, 647).

Tien-Cuong Dinh and Keiji Oguiso (2017). "A surface with discrete and non-finitely generated automorphism group". arXiv: 1710.07019 (cit. on p. 642).

Tien-Cuong Dinh and Nessim Sibony (2005). "Une borne supérieure pour l'entropie topologique d'une application rationnelle". *Ann. of Math. (2)* 161.3, pp. 1637–1644. MR: 2180409 (cit. on pp. 645, 647).

– (2010a). "Dynamics in several complex variables: endomorphisms of projective spaces and polynomial-like mappings". In: *Holomorphic dynamical systems*. Vol. 1998. Lecture Notes in Math. Springer, Berlin, pp. 165–294. MR: 2648690 (cit. on p. 646).

– (2010b). "Super-potentials for currents on compact Kähler manifolds and dynamics of automorphisms". *J. Algebraic Geom.* 19.3, pp. 473–529. MR: 2629598 (cit. on p. 650).

Romain Dujardin (2006). "Laminar currents and birational dynamics". *Duke Math. J.* 131.2, pp. 219–247. MR: MR2219241 (cit. on p. 649).

Benson Farb and Dan Margalit (2012). *A primer on mapping class groups*. Vol. 49. Princeton Mathematical Series. Princeton, NJ: Princeton University Press, pp. xiv+472. MR: 2850125 (cit. on pp. 643, 650).

Harry Furstenberg (1961). "Strict ergodicity and transformation of the torus". *Amer. J. Math.* 83, pp. 573–601. MR: 0133429 (cit. on p. 648).

Marat H. Gizatullin (1980). "Rational G-surfaces". *Izv. Akad. Nauk SSSR Ser. Mat.* 44.1, pp. 110–144, 239. MR: 563788 (cit. on pp. 646, 647).

Phillip Griffiths and Joseph Harris (1978). *Principles of algebraic geometry*. Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York, pp. xii+813. MR: 507725 (cit. on p. 642).

Mikhaïl Gromov (1987). "Entropy, homology and semialgebraic geometry". *Astérisque* 145-146. Séminaire Bourbaki, Vol. 1985/86, pp. 5, 225–240. MR: 880035 (cit. on p. 645).

– (2003). "On the entropy of holomorphic maps". *Enseign. Math. (2)* 49.3-4, pp. 217–235. MR: 2026895 (cit. on p. 645).

Alexander Grothendieck (1995). "Techniques de construction et théorèmes d'existence en géométrie algébrique. IV. Les schémas de Hilbert". In: *Séminaire Bourbaki, Vol. 6*. Soc. Math. France, Paris, Exp. No. 221, 249–276. MR: 1611822 (cit. on p. 640).

János Kollár and Frédéric Mangolte (2009). "Cremona transformations and diffeomorphisms of surfaces". *Adv. Math.* 222.1, pp. 44–61. MR: 2531367 (cit. on p. 644).

Vladimir Lazić, Keiji Oguiso, and Thomas Peternell (2016). "Nef line bundles on Calabi-Yau threefolds, I". arXiv: 1601.01273 (cit. on p. 647).

John Lesieutre (2017). "A projective variety with discrete, non-finitely generated automorphism group". *Inventiones Math., to appear*, pp. 1–13 (cit. on p. 642).

David I. Lieberman (1978). "Compactness of the Chow scheme: applications to automorphisms and deformations of Kähler manifolds". In: *Fonctions de plusieurs variables complexes, III (Sém. François Norguet, 1975–1977)*. Vol. 670. Lecture Notes in Math. Springer, Berlin, pp. 140–186. MR: 521918 (cit. on p. 640).

Jan-Li Lin (2013). "On degree growth and stabilization of three-dimensional monomial maps". *Michigan Math. J.* 62.3, pp. 567–579. MR: 3102531 (cit. on p. 647).

Federico Lo Bianco (2014). "Bornes sur les degrés dynamiques d'automorphismes de variétés kählériennes de dimension 3". *C. R. Math. Acad. Sci. Paris* 352.6, pp. 515–519. MR: 3210135 (cit. on pp. 645, 646).

Curtis T. McMullen (2002a). "Coxeter groups, Salem numbers and the Hilbert metric". *Publ. Math. Inst. Hautes Études Sci.* 95, pp. 151–183. MR: 1953192 (cit. on p. 645).

– (2002b). "Dynamics on $K3$ surfaces: Salem numbers and Siegel disks". *J. Reine Angew. Math.* 545, pp. 201–233. MR: 1896103 (cit. on p. 649).

– (2007). "Dynamics on blowups of the projective plane". *Publ. Math. Inst. Hautes Études Sci.* 105, pp. 49–89. MR: 2354205 (cit. on p. 645).

Arnaud Moncet (2008). "Un exemple de groupe d'automorphismes d'une surface K3 réelle" (cit. on p. 644).

– (2012). "Real versus complex volumes on real algebraic surfaces". *Int. Math. Res. Not. IMRN* 16, pp. 3723–3762. MR: 2959025 (cit. on p. 650).

Tuyen Trung Truong (2016). "Relative dynamical degrees of correspondences over a field of arbitrary characteristic". arXiv: 1701.07760 (cit. on pp. 645, 647).

Christian Urech (2017). "Remarks on the degree growth of birational transformations". *Math. Research Letters, to appear*, pp. 1–12 (cit. on p. 647).

Jörg Winkelmann (2004). "Realizing connected Lie groups as automorphism groups of complex manifolds". *Comment. Math. Helv.* 79.2, pp. 285–299. MR: 2059433 (cit. on p. 641).

Yosef Yomdin (1987). "Volume growth and entropy". *Israel J. Math.* 57.3, pp. 285–300. MR: 889979 (cit. on p. 645).

De-Qi Zhang (2013). "Algebraic varieties with automorphism groups of maximal rank". *Math. Ann.* 355.1, pp. 131–146. MR: 3004578 (cit. on p. 643).

Received 2017-11-29.

Serge Cantat
Univ Rennes
CNRS, IRMAR - UMR 6625
F-35000 Rennes
France
serge.cantat@univ-rennes1.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (653–670)

RECURSIVE COMBINATORIAL ASPECTS OF COMPACTIFIED MODULI SPACES

Lucia Caporaso

Abstract

In recent years an interesting connection has been established between some moduli spaces of algebro-geometric objects (e.g. algebraic stable curves) and some moduli spaces of polyhedral objects (e.g. tropical curves).

In loose words, this connection expresses the Berkovich skeleton of a given algebro-geometric moduli space as the moduli space of the skeleta of the objects parametrized by the given space; it has been proved to hold in two important cases: the moduli space of stable curves and the moduli space of admissible covers. Partial results are known in other cases.

This connection relies on the study of the boundary of the algebro-geometric moduli spaces and on its recursive, combinatorial properties, some of which have been long known and are now viewed from a new perspective.

Contents

1 Introduction

We will describe some results in this area by focusing on the moduli spaces of curves, line bundles on curves (i.e. Jacobians), and coverings of curves. As we said, it was clear

MSC2010: primary 14H10; secondary 14TXX.

for a long time that combinatorial aspects play a significant role in compactifying moduli spaces. A notable example is the structure of the Néron model of the Jacobian of a curve. Of course, Néron models are not compact, but they are a first step towards compactifying Jacobians. We begin the paper by describing Néron models, their combinatorial properties, and the recursive structure of their compactification. Then we turn to moduli spaces of curves and illustrate the connection introduced at the beginning. We return to compactified Jacobians in the last section and describe some recent partial results.

2 Compactified Jacobians and Néron models

2.1 Jacobians, Picard schemes and Néron models. Let X be a connected, reduced, projective curve of genus g over an algebraically closed field k. The set of isomorphism classes of line bundles of degree 0 on every irreducible component of X is an algebraic group of dimension g, denoted by J_X, and called the Jacobian of X. More exactly, J_X is the moduli space for line bundles of multidegree $(0, 0, \dots, 0)$ on X.

J_X is projective, hence an abelian variety, if X is smooth but not if X is singular (with exceptions). Constructing compactifications for J_X is a classical problem which can be stated as follows. Let

$$X \hookrightarrow \mathcal{X} \xrightarrow{f} \operatorname{Spec} R$$

be a family of curves over the spectrum of a discrete valuation ring R, with X as special fiber. Assume the generic fiber, $\mathcal{X}_K$, to be a nonsingular curve over the quotient field, K, of R, and let $J_{\mathcal{X}_K} \to \operatorname{Spec} K$ be its Jacobian. The problem is to find "good" models of $J_{\mathcal{X}_K}$ over Spec R, where "good" means: (a) projective, (b) with a moduli interpretation, (c) such that the special fiber depends only on X and not on the family f. We shall refer to such models, and to their special fiber, as *compactified Jacobians*.

We shall begin by looking for models that satisfy (b) and (c) together with the weaker requirement of being separated, rather than projective.

A natural model of $J_{\mathcal{X}_K}$ is the relative Jacobian (a group scheme over R)

$$J_{\mathcal{X}/R} \longrightarrow \operatorname{Spec} R. \tag{1}$$

This is a separated model for $J_{\mathcal{X}_K}$, but it does not have satisfactory moduli properties. Indeed, suppose we have a line bundle $\mathcal{L}$ over $\mathcal{X}$ having relative degree 0 on the fibers of f. Then we have a *moduli morphism* $\mu_{\mathcal{L}_K} : \operatorname{Spec} K \to J_{\mathcal{X}_K}$ whose image corresponds to the restriction of $\mathcal{L}$ to $\mathcal{X}_K$. For a model to have good moduli properties we want the map $\mu_{\mathcal{L}_K}$ to extend to a morphism from Spec R to the model (so that the image of the special point is determined by the restriction of $\mathcal{L}$ to the special fiber). For the relative Jacobian this requirement easily fails.

Another natural model is the relative degree-0 Picard scheme

$$\mathrm{Pic}^0_{\mathfrak{X}/R} \longrightarrow \mathrm{Spec}\, R. \tag{2}$$

Now, this has a good moduli interpretation, as any moduli map $\mu_{\mathfrak{L}_K}$ will certainly extend to a map $\mu_R : \mathrm{Spec}\, R \to \mathrm{Pic}^0_{\mathfrak{X}/R}$, and μ_R is itself a moduli map. The problem now is that the extension μ_R may fail to be unique, that is, (2) is not a separated morphism, in general.

We have thus two natural models, one separated with bad moduli properties, the other with good moduli properties but not separated. Does there exist a compromise with better behaviour than both of them? An answer to this question comes from the theory of Néron models.

Our moduli requirement above (i.e. the existence of a unique extension for maps of the form $\mu_{\mathfrak{L}_K}$) is a special case of the mapping property satisfied by Néron models, whose existence has been established in Néron [1964]. We state the following special case of Néron's famous theorem, using the above notation.

Theorem 2.1.1 (Néron)**.** *There exists a smooth and separated group scheme of finite type*

$$N(J_{\mathfrak{X}_K}) \longrightarrow \mathrm{Spec}\, R,$$

whose generic fiber is $J_{\mathfrak{X}_K}$, satisfying the following mapping property:

Let $Y_R \to \mathrm{Spec}\, R$ be smooth, Y_K its generic fiber, and $\mu_K : Y_K \to J_{\mathfrak{X}_K}$ a morphism. Then μ_K extends uniquely to a morphism $\mu_R : Y_R \to N(J_{\mathfrak{X}_K})$.

Remark 2.1.2. The Néron model does not commute with ramified base change, therefore it is a separated model for $J_{\mathfrak{X}_R}$ with good mapping properties, but not functorial ones.

The relation of the Néron model with the relative Jacobian and the relative Picard scheme has ben established by Raynaud who proved, in Raynaud [1970], that $N(J_{\mathfrak{X}_K}) \to \mathrm{Spec}\, R$ is the maximal separated quotient of $\mathrm{Pic}^0_{\mathfrak{X}/R} \to \mathrm{Spec}\, R$.

2.2 Combinatorics of Néron models. We shall turn to the geometric structure of our Néron models. We assume from now on that the curve X is nodal, write $X = \cup_{v\in V} C_v$ for its irreducible components, and consider the dual graph, G_X, of X:

$$G_X := (V, E = \mathrm{Sing}(X), V \xrightarrow{h} \mathbb{Z}) \tag{3}$$

with $h(v) = g(C_v)$, where $g(C_v)$ is the genus of the normalization, C_v^ν. The genus of G_X is the same as the arithmetic genus of X, i.e.

$$g(X) = g(G_X) = b_1(G_X) + \sum_{v\in V} h(v).$$

It is well known that J_X fits into an exact sequence of algebraic groups:

$$0 \longrightarrow (k^*)^{b_1(G_X)} \longrightarrow J_X \longrightarrow \prod_{v \in V} J_{C_v^\nu} \longrightarrow 0. \tag{4}$$

Denote by N_X the special fiber of $N(J_{\mathcal{X}_K}) \to \operatorname{Spec} R$. We shall state some results of Raynaud [1970], and Oda and Seshadri [1979], which establish that N_X is a disjoint union of copies of the Jacobian of X indexed by a combinatorial invariant of the curve.

To do that we need some combinatorial preliminaries. Fix an orientation on $G = G_X$ (whose choice is irrelevant), let $C_0(G,\mathbb{Z})$ and $C_1(G,\mathbb{Z})$ be the standard groups of i-chains, generated over $\mathbb{Z}$ by V if $i = 0$, and by E if $i = 1$. Next, let $\partial : C_1(G,\mathbb{Z}) \to C_0(G,\mathbb{Z})$ be the usual boundary (mapping an edge e oriented from u to v to $u - v$), and $\delta : C_0(G,\mathbb{Z}) \to C_1(G,\mathbb{Z})$ the coboundary (mapping a vertex v to $\sum e_v^+ - \sum e_v^-$ where the first sum is over all edges originating from v and the second over all edges ending at v).

Now we can state the following.

Proposition 2.2.1 (Raynaud, Oda-Seshadri)**.** *Let $\mathcal{X} \to \operatorname{Spec} R$ have nodal special fiber and regular total space. Then the special fiber, N_X, of the Néron model $N(J_{\mathcal{X}_K}) \longrightarrow \operatorname{Spec} R$ is a union of copies of J_X as follows*

$$N_X \cong \sqcup_{i \in \Phi_X} (J_X)_i \tag{5}$$

where Φ_X is the following finite group

$$\Phi_G \cong \frac{\partial\delta C_0(G,\mathbb{Z})}{\partial C_1(G\mathbb{Z})}.$$

In particular, the number of irreducible components of N_X equals the number of spanning trees of G_X.

The last claim follows from Kirchhoff-Trent, or Kirchhoff matrix,Theorem.

2.3 Compactifying Néron models. From now on we apply the notation introduced in Proposition 2.2.1 and for any connected nodal curve X we denote by N_X the special fiber of the Néron model of the Jacobian associated to a family $\mathcal{X} \to \operatorname{Spec} R$ with $\mathcal{X}$ regular.

If X is a singular curve with $b_1(G_X) \neq 0$ (i.e. X not of "compact type"), then the Néron model $N(J_{\mathcal{X}_K}) \to \operatorname{Spec} R$ is not projective, as its special fiber is not projective by the exact sequence (4).

Now, the Picard scheme has good moduli properties and the Néron model is its maximal separated quotient. We introduce a terminology to distinguish compactified Jacobians which also compactify the Néron model.

A compactified Jacobian $\overline{P} \longrightarrow \operatorname{Spec} R$ is said to be of *Néron type* (or a *Néron compactified Jacobian*) if its special fiber, written $\overline{P}_X$, contains N_X as a dense open subset. We shall also say that $\overline{P}_X$ is of Néron type.

As we shall see, Néron compactified Jacobians do exist, but there exist also interesting compactified Jacobians not of Néron type.

The notion originates from Oda and Seshadri [ibid.] and Caporaso [2008], although the terminology was introduced later, in Caporaso [2012b]. Oda and Seshadri, in Oda and Seshadri [1979], treated the case of a fixed singular curve X (rather than a family of curves), and constructed compactified Jacobians in this less general sense. They nonetheless established the link with Néron models and constructed compactified Jacobians whose smooth locus is isomorphic to N_X. This was extended to families of curves later. In Caporaso [1994] and Caporaso [2008] a class of compactified Jacobians of Néron type was proved to exist and to form a family over the moduli space of stable curves, $\overline{\mathfrak{M}}_g$. Such families, denoted by $\psi_d : \overline{\mathcal{P}}_g^d \longrightarrow \overline{\mathfrak{M}}_g$, are indexed by the integers d such that

$$(d - g + 1, 2g - 2) = 1. \tag{6}$$

For any stable curve X the fiber of ψ_d over X, written $\overline{P}_X^d$, contains N_X as a dense open subset equal to its smooth locus, and it is thus of Néron type.

In a similar vein, in Melo and Viviani [2012] and in Melo, Rapagnetta, and Viviani [2017] other Néron compactified Jacobians were found among the ones constructed by Esteves [2001], and called "fine" compactified Jacobians. The word "fine" is quite appropriate, as all known Néron compactified Jacobians are as fine a moduli space as they can be, i.e. they admit a universal (or "Poincaré") line bundle.

We now show how Néron Jacobians are recursive compactifications of Néron models. First, for a connected graph G, we introduce the following

$$\mathcal{C}(G) := \{S \subset E(G) : \; G - S \text{ is connected}\},$$

with partial order given by reverse inclusion. The maximal element of $\mathcal{C}(G)$ is $\emptyset$, and the minimal elements are the $S \subset E$ such that $G - S$ is a spanning tree. Moreover, $\mathcal{C}(G)$ is a graded poset with respect to the rank function $S \mapsto g(G - S)$.

If G is the dual graph of the curve X then $S \in \mathcal{C}(G)$ is a set of nodes of X. We denote by X_S^ν the desingularization of X at S, so that X_S^ν is a connected nodal curve of genus $g(G - S)$. Recall that $N_{X_S^\nu}$ denotes the special fiber of the Néron model of its Jacobian. The following follows from Caporaso [2008, Thm. 7.9].

Theorem 2.3.1. *Let $\overline{P}_X^d$ be a Néron compactified Jacobian. Then*

$$\overline{P}_X^d = \bigsqcup_{S \in \mathcal{C}(G)} N_S \tag{7}$$

with $N_S \cong N_{X_S^\nu}$ *for every* $S \in \mathcal{C}(G)$*. Moreover* (7) *is a* graded stratification*, i.e. the following hold.*

1. $N_S \cap \overline{N_{S'}} \neq \emptyset \Leftrightarrow N_S \subset \overline{N_{S'}} \Leftrightarrow S' \geq S$.
2. N_S *is locally closed of pure dimension* $g(G - S)$.
3. *The following is a rank-function on* $\mathcal{C}(G)$

$$\mathcal{C}(G) \longrightarrow \mathbb{Z}; \qquad S \mapsto \dim N_S.$$

Remark 2.3.2. The set of strata of minimal dimension in (7) are Néron models of curves whose dual graph is a spanning tree of G_X, hence they are irreducible. By Proposition 2.2.1, the number of such strata is equal to the number of irreducible components of $\overline{P}^d_X$.

If X is not of compact type the strata of (7) are not all connected. Hence one naturally asks whether the stratification can be refined so as to have connected strata. We will answer this question later in the paper.

Remark 2.3.3. The theorem exhibits the compactification of the Néron model of X in terms of the Néron models of partial normalizations of X. This phenomenon is an instance of what seems to be a widespread recursive behaviour for compactified moduli spaces. Namely, to compactify a space (e.g. N_X) one adds at the boundary the analogous spaces associated to simpler objects (e.g. $N_{X'}$ with X' a connected partial normalization of X). Other examples of this recursive pattern will appear later in the paper.

The concept of "graded stratification" used in the Theorem will appear again, so we now define it in general.

A *graded stratification* of an algebraic variety, or a stack, M by a poset $\mathcal{P}$ is a partition $M = \bigsqcup_{p \in \mathcal{P}} M_p$ such that the following hold for every $p, p' \in \mathcal{P}$.

1. $M_p \cap \overline{M_{p'}} \neq \emptyset \Leftrightarrow M_p \subset \overline{M_{p'}} \Leftrightarrow p' \geq p$.
2. M_p is equidimensional and locally closed.
3. The map from $\mathcal{P}$ to $\mathbb{N}$ sending p to $\dim M_p$ is a rank function on $\mathcal{P}$.

3 Moduli of curves: tropicalization and analytification

3.1 Moduli spaces of algebraic and tropical curves. Let $\mathfrak{M}_{g,n}$ be the moduli space of smooth curves of genus g with n marked points. Assume $2g - 2 + n > 0$ so that $\mathfrak{M}_{g,n}$

is never empty. It is well known that $\mathfrak{M}_{g,n}$ is not projective (unless $g = 0, n = 3$) and is compactified by the moduli space of Deligne-Mumford stable curves, $\overline{\mathfrak{M}}_{g,n}$; see Deligne and Mumford [1969], Knudsen [1983a,b], and Gieseker [1982].

We will describe in (9) a stratification of $\overline{\mathfrak{M}}_{g,n}$ which is recursive in the sense of Remark 2.3.3, that is, the boundary strata are expressed in terms of simpler moduli spaces $\mathfrak{M}_{g',n'}$. In this case "simpler"means $g \leq g'$ and of smaller dimension, i.e. $n' < n+3(g-g')$. These boundary strata will be described in (8).

Let $\mathcal{SG}_{g,n}$ be the poset of stable graphs of genus g with n legs, with the following partial order:

$$G_2 \geq G_1 \quad \text{if} \quad G_2 = G_1/S \quad \text{for some} \quad S \subset E(G_1)$$

where G_1/S is the graph obtained by contracting every edge of S to a vertex. By the very definition (introduced in Brannetti, Melo, and Viviani [2011]), edge-contraction preserves g and n, and it is easily seen to preserve stability.

To a stable curve, X, of genus g with n legs there corresponds a dual graph $G_X \in \mathcal{SG}_{g,n}$. With respect to what we defined in (3) the only new piece of data are the legs of G_X which correspond to the n marked points. Recall that the weight, $h(v)$, of a vertex v is the geometric genus of the corresponding component of X. We denote by $\deg(v)$ the degree (or valency) of v.

For $G \in \mathcal{SG}_{g,n}$ we denote by $\mathfrak{M}_G$ the moduli stack of curves whose dual graph is isomorphic to G. We have (see Abramovich, Caporaso, and Payne [2015, Prop. 3.4.1])

$$\mathfrak{M}_G = \Big[\Big(\prod_{v \in V(G)} \mathfrak{M}_{h(v),\deg(v)}\Big)/\mathrm{Aut}(G)\Big]. \tag{8}$$

With this notation and the terminology at the end of Section 2 we state:

Proposition 3.1.1. *The following is a graded stratification of* $\overline{\mathfrak{M}}_{g,n}$

$$\overline{\mathfrak{M}}_{g,n} = \bigsqcup_{G \in \mathcal{SG}_g} \mathfrak{M}_G. \tag{9}$$

One goal of the above descriptive result (whose proof, in Caporaso [2012a], is not hard thanks to our consolidated knowledge of $\overline{\mathfrak{M}}_g$) is to highlight the similarities of $\overline{\mathfrak{M}}_{g,n}$ with the moduli space of extended (abstract) tropical curves, $\overline{M}^{\mathrm{trop}}_{g,n}$, as we are going to show. First of all, an extended tropical curve is a metric graph, i.e. a graph G whose edges are assigned a length, $\ell : E \to \mathbb{R}_{>0} \cup \{\infty\}$. We denote a tropical curve as follows

$$\Gamma = (G; \ell) = (V, E, w; \ell).$$

The genus of Γ is the genus of G. The word "extended" refers to the fact that we include edges of infinite length. In fact, abstract tropical curves are originally defined (in Mikhalkin and Zharkov [2008] and Brannetti, Melo, and Viviani [2011]) as compact spaces, so their edges have finite length. We allow edges of infinite lengths to obtain a compact moduli space which is the "tropicalization" of the moduli space of stable curves.

An extended tropical curve with n marked points is a metric graph as above with the addition of a set of legs of the underlying graph.

There is a natural equivalence relation on tropical curves such that in every equivalence class there is a unique (up to isomorphism) curve whose underlying graph is stable.

The moduli space $\overline{M}^{\rm trop}_{g,n}$ parametrizes extended tropical curves of genus g with n marked points up to this equivalence relation. Its first construction is due to Mikhalkin, for $g = 0$, and to Brannetti-Melo-Viviani for compact tropical curves in any genus. The following statement summarises results from Mikhalkin [2007], Brannetti, Melo, and Viviani [2011], and Caporaso [2012a].

Theorem 3.1.2. *The moduli space of extended tropical curves, $\overline{M}^{\rm trop}_{g,n}$, is a compact and normal topological space of dimension $3g-3+n$. It admits a graded stratification*

$$\overline{M}^{\rm trop}_{g,n} = \bigsqcup_{G \in \mathcal{SG}^*_{g,n}} \overline{M}^{\rm trop}_{G} \tag{10}$$

where $\overline{M}^{\rm trop}_{G}$ is the locus of curves having G as underlying graph and $\mathcal{SG}^*_{g,n}$ is the poset dual to $\mathcal{SG}_{g,n}$ (i.e. with reverse partial order).

3.2 Skeleta and tropicalizations. Proposition 3.1.1 and Theorem 3.1.2 show that $\overline{M}^{\rm trop}_{g,n}$ has a graded stratification dual to that of $\overline{\mathfrak{M}}_{g,n}$. Therefore we ask whether there exists some deeper relation between $\overline{\mathfrak{M}}_{g,n}$ and $\overline{M}^{\rm trop}_{g,n}$.

A positive answer can be given through the theory of analytifications of algebraic schemes developed in Berkovich [1990], and through its connections to tropical geometry; see Maclagan and Sturmfels [2015] and Payne [2009]. Let us introduce the space $\overline{M}^{\rm an}_{g,n}$, the analytification of $\overline{\mathfrak{M}}_{g,n}$ in the sense of Berkovich. Recall that a point in $\overline{M}^{\rm an}_{g,n}$ corresponds, up to base change, to a stable curve over an algebraically closed field K complete with respect to a non-archimedean valuation; as $\overline{M}_{g,n}$ is projective, this is the same as a stable curve over the ring of integers of K.

From the general theory (see also Thuillier [2007] and Abramovich, Caporaso, and Payne [2015]) we have that for every space with a toroidal structure, like the stack $\overline{\mathfrak{M}}_{g,n}$ (with toroidal structure given by its boundary), one associates the *Berkovich skeleton* which is a generalised, extended cone complex onto which the analytification retracts. We write $\overline{\Sigma}(\overline{\mathfrak{M}}_{g,n})$ for the Berkovich skeleton of $\overline{\mathfrak{M}}_{g,n}$ and $\overline{M}^{\rm an}_{g,n} \xrightarrow{\rho} \overline{\Sigma}(\overline{\mathfrak{M}}_{g,n})$ for the retraction.

The connection between tropical geometry and Berkovich theory originates from the fact that Berkovich skeleta can be viewed as tropicalizations of algebraic varieties. As we are going to see, the picture for curves is quite clear, whereas other interesting situations (some treated later in the present paper) are still open to investigations; we refer to Gubler, Rabinoff, and Werner [2016], for recent progress for higher dimensional varieties.

Now, tropical curves can be viewed as tropicalizations (or skeleta) of curves over algebraically closed, complete, non-archimedean fields. This is clarified by the following statement, combining results of Viviani [2013], Tyomkin [2012], and Baker, Payne, and Rabinoff [2016, 2013] (for the first part) and of Abramovich, Caporaso, and Payne [2015] (for the second part).

Theorem 3.2.1. *There exists a* tropicalization *map*

$$\mathrm{trop}: \overline{M}^{\mathrm{an}}_{g,n} \to \overline{M}^{\mathrm{trop}}_{g,n}$$

that sends the class of a stable curve over the (algebraically closed, complete, non-archimedean) field K to its skeleton.

There is a natural isomorphism, $\overline{\Sigma}(\overline{\mathfrak{M}}_{g,n}) \cong \overline{M}^{\mathrm{trop}}_{g,n}$, through which the above map factors as follows

$$\mathrm{trop}: \overline{M}^{\mathrm{an}}_{g,n} \xrightarrow{\rho} \overline{\Sigma}(\overline{\mathfrak{M}}_{g,n}) \xrightarrow{\cong} \overline{M}^{\mathrm{trop}}_{g,n}.$$

We need to define the tropicalization map and explain the word "skeleton". As we said, a point in $\overline{M}^{\mathrm{an}}_{g,n}$ is a class of stable curves, $\mathfrak{X} \to \operatorname{Spec} R$, where R is the valuation ring of K; let X be its special fiber. Then the image of this point via the map trop is the tropical curve $(G;\ell)$, where G is the dual graph of X and, for every node $e \in E(G)$, the value $\ell(e)$ is determined by the local geometry of $\mathfrak{X}$ at e as measured by the given valuation. Such a tropical curve $(G;\ell)$ is called the *skeleton*, of the stable curve $\mathfrak{X} \to \operatorname{Spec} R$ or of the stable curve $\mathfrak{X}_K$ over K.

Concluding in loose words: the skeleton of $\overline{\mathfrak{M}}_{g,n}$ is the moduli space of skeleta of stable curves.

The structure of the proof of the above theorem is such that it may apply to other situations. In fact it has been applied by Cavalieri-Markwig-Ranganathan to another remarkable case, the compactification of the Hurwitz spaces, as we shall now explain.

3.3 Algebraic and tropical admissible covers. Consider the moduli space of admissible covers, $\overline{\mathcal{H}}_\bullet$, which here (as in Cavalieri, Markwig, and Ranganathan [2016]), is the normal compactification of the classical space of Hurwitz covers $\mathcal{H}_\bullet$. We use the subscript •" to simplify the notation needed to express the usual discrete invariants. Indeed, a more precise notation would be $\mathcal{H}_\bullet = \mathcal{H}_{g,h}(\underline{\pi})$ for the Hurwitz space parametrizing degree-d covers of a smooth curve of genus h by a smooth curve of genus g with exactly

b branch points with ramification profile prescribed by a set of b partitions of d, written $\underline{\pi} = (\pi_1, \ldots, \pi_b)$.

In Cavalieri, Markwig, and Ranganathan [2016] the authors define *tropical admissible covers*, construct their moduli space $\overline{H}_\bullet^{\text{trop}}$ and establish an analogue to Theorem 3.2.1. We shall now outline the procedure and give some details.

The first step is to associate a dual combinatorial entity to the algebro-geometric one. Indeed, one associates to an admissible cover a map of graphs, which we call the *dual graph cover*. The set of all dual graph covers is endowed with a poset structure by means of edge-contractions (similarly to the poset set of stable graphs). We denote by $\mathcal{A}_\bullet$ this poset, as its objects can be viewed as *admissible* maps of graphs. For any $\Theta \in \mathcal{A}_\bullet$ we denote by $\mathcal{H}_\Theta$ the locus in $\overline{\mathcal{H}}_\bullet$ of admissible covers whose dual graph map is Θ.

The second step is to enrich the dual combinatorial entity with a tropical, or polyhedral, structure. In Cavalieri, Markwig, and Ranganathan [ibid.] tropical admissible covers are defined by metrizing, in a suitable way, dual maps of graphs. The moduli space of tropical admissible covers is denoted by $\overline{H}_\bullet^{\text{trop}}$, the bar over H indicates that they are "extended", i.e. edge-lengths are allowed to be infinite. For $\Theta \in \mathcal{A}_\bullet$ the stratum parametrizing tropical admissible covers having Θ as underlying graph map is denoted by $\overline{H}_\Theta^{\text{trop}}$ and shown to be the quotient of an extended real cone.

The third and last step is to use analytification and tropicalization to establish an explicit link between the algebraic and the tropical moduli space. Indeed, essentially by construction, we have dual stratifications

$$\overline{\mathcal{H}}_\bullet = \bigsqcup_{\Theta \in \mathcal{A}_\bullet} \mathcal{H}_\Theta \qquad \text{and} \qquad \overline{H}_\bullet^{\text{trop}} = \bigsqcup_{\Theta \in \mathcal{A}_\bullet^*} \overline{H}_\Theta^{\text{trop}}.$$

This duality can be read from the following Theorem.

Theorem 3.3.1 (Cavalieri-Markwig-Ranganathan)**.** *There is a tropicalization map* trop : $\overline{\mathcal{H}}_\bullet^{\text{an}} \to \overline{H}_\bullet^{\text{trop}}$ *which factors as follows*

$$\text{trop} : \overline{\mathcal{H}}_\bullet^{\text{an}} \xrightarrow{\rho} \overline{\Sigma}(\overline{\mathcal{H}}_\bullet) \longrightarrow \overline{H}_\bullet^{\text{trop}}.$$

The map $\overline{\mathcal{H}}_\bullet^{\text{an}} \xrightarrow{\rho} \overline{\Sigma}(\overline{\mathcal{H}}_\bullet)$ is the retraction of $\overline{\mathcal{H}}_\bullet^{\text{an}}$ onto its Berkovich skeleton, as described in subSection 3.2. This result is compatible with the analogous one for $\overline{\mathcal{M}}_g$ through the canonical forgetful maps from $\overline{\mathcal{H}}_\bullet$ to the moduli spaces of stable curves; see Cavalieri, Markwig, and Ranganathan [ibid., Thm. 4].

What about other moduli spaces? Are Theorems 3.2.1 and 3.3.1 part of some general picture where skeleta of algebraic moduli spaces (e.g. the skeleton of $\overline{\mathcal{M}}_{g,n}$) can be described as moduli spaces for combinatorial entities (e.g. tropical curves) which are skeleta of the objects (e.g. stable curves) parametrized by the algebraic moduli spaces? As we saw, the first step is to identify a suitable partially ordered set of combinatorial objects to associate to the algebro-geometric ones.

In the next section we shall look at the theory of compactified Jacobians from this point of view.

4 Compactified Jacobians

4.1 Compactifying Jacobians over $\overline{M}_g$. Let us go back to compactify Jacobians of curves and, with the discussion of the previous section in mind, approach the problem from the point of view of the moduli theory of stable curves. Consider the universal Jacobian over the moduli space of smooth curves and look for a compactification of it over $\overline{\mathfrak{M}}_g$ satisfying the requirements we discussed earlier.

For reasons that will be clear later, it is convenient to extend our considerations to Jacobians of all degree. For a curve $X = \cup_{v \in V} C_v$ and any multidegree $\underline{d} \in \mathbb{Z}^V$ we write $\mathrm{Pic}^{\underline{d}}(X)$ for the moduli space of line bundles of multidegree $\underline{d}$. Now, J_X is identified with $\mathrm{Pic}^{(0,\ldots,0)}(X)$ and we have non-canonical isomorphisms $J_X \cong \mathrm{Pic}^{\underline{d}}(X)$.

The universal degree-d Jacobian over $\mathfrak{M}_g$ is a morphism $\mathcal{P}_g^d \to \mathfrak{M}_g$, whose fiber over the point parametrizing a smooth curve X is $\mathrm{Pic}^d(X)$. We want to construct a compactification of $\mathcal{P}_g^d$ over $\overline{\mathfrak{M}}_g$ by a projective morphism $\overline{\mathcal{P}}_g^d \to \overline{\mathfrak{M}}_g$ such that $\overline{\mathcal{P}}_g^d$ has a moduli description. We shall refer to such a space as a compactified universal degree-d Jacobian.

From this perspective there is a natural approach to the problem, namely imitate the construction of $\overline{\mathfrak{M}}_g$ itself. Recall that the moduli scheme, $\overline{M}_g$, of the stack $\overline{\mathfrak{M}}_g$ was constructed by Gieseker [1982] as the GIT-quotient of the Hilbert scheme of n-canonically embedded curves ($n \gg 0$). The moduli stack $\overline{\mathfrak{M}}_g$ is the quotient stack associated to this quotient.

Now, as the Hilbert scheme of n-canonically embedded curves of genus g has such a beautiful GIT-quotient, why shouldn't the Hilbert scheme of all projective curves of degree $d \gg 0$ and genus g have a beautiful GIT-quotient? And why shouldn't this quotient be a candidate for a compactification of the universal degree-d Jacobian? Indeed, this is what happens, and the GIT quotient of this Hilbert scheme is our compactified universal degree-d Jacobian, $\overline{P}_g^d$. The corresponding quotient stack is denoted by $\overline{\mathcal{P}}_g^d$.

Now, as d varies, the spaces $\overline{\mathcal{P}}_g^d$ are not isomorphic to one another, and their fibers over certain singular curves are not even birational to one another. Again, we see the phenomenon (appearing already in Oda and Seshadri [1979]) that non isomorphic compactifications of the Jacobian of a fixed singular curve exist. In the present case the various models depend on the degree d.

As for the basic properties of the spaces $\overline{\mathcal{P}}_g^d$ as d varies, there is a special set of degrees, namely those such that (6) holds, such that $\overline{P}_g^d$ is a geometric GIT-quotient and has good moduli properties, so that its points correspond to geometric objects up to a certain equivalence relation. Moreover, the natural (projective) morphism $\psi_d : \overline{\mathcal{P}}_g^d \longrightarrow \overline{\mathfrak{M}}_g$ is a strongly representable map of Deligne-Mumford stacks. As we said in subSection 2.3 in this case all fibers of ψ_d are Néron compactified Jacobians.

Now, (6) fails if $d = g - 1$ whereas it holds if $d = g$. We shall concentrate on these two cases, interesting for different reasons, and give a combinatorial analysis of the compactified Jacobian.

4.2 Compactified Jacobians in degree $g-1$.

We begin by reviewing an idea of Beauville. Let us fix $g \geq 2$. Among degree-d Jacobians, the case $d = g - 1$ has been object of special interest for its strong connections with the Theta divisor, the Schottky problem, the Prym varieties; in particular, it has been studied in Beauville [1977].

Let us approach the problem of compactifying the degree $(g-1)$-Jacobian of a curve X with $G = (V, E, h)$ as dual graph. We expect a good compactification to have finitely many irreducible components and each component to parametrize (at least away from the boundary) line bundles on X of a fixed multidegree $\underline{d}$ such that $|\underline{d}| = g - 1$. Now the question is to determine these “special” multidegrees. Consider the identity

$$g - 1 = \sum_{v \in V} \big(h(v) - 1\big) + |E|.$$

We can interpret the first term (i.e. the summation) as carrying the topological invariants of X, and the second term, $|E|$, as carrying the combinatorial ones. Now, while the first term exhibits the single contribution of each vertex/component, the second does not. So we may ask how to distribute the second term among the various vertices in a combinatorially meaningful way. A natural solution is to consider an orientation, O, on G, and denote by $\underline{t}_v^O$ the number of edges having v as target. Then $\sum_{v\in V} \underline{t}_v^O = |E|$. Therefore, if we define a multidegree $\underline{d}^O$ as follows

$$\underline{d}_v^O := h(v) - 1 + \underline{t}_v^O,$$

for every $v \in V$, we have $|\underline{d}^O| = g - 1$. As there are only finitely many orientations on a graph, by the above rule we have picked a finite set of special multidegrees of degree $g - 1$. More precisely, it may happen that two orientations, O and O', give the same multidedegree; in such a case we say that O and O' are *equivalent*. We denote by $\overline{O}(G)$ the set of such equivalence classes of orientations on G. Now, a closer look reveals that $\overline{O}(G)$ is still too big for it to parametrize the irreducible components of a compactified

Jacobian. Indeed, from the discussion in SubSection 2.2 we expect the number of such components to be at most equal to $|\Phi_G|$, whereas we have $|\overline{O}(G)| > |\Phi_G|$ in general (for example for a 2-cycle, as in the picture below).

So, to compactify the degree-$(g-1)$ Jacobian we must distinguish a special type of orientations. These are called *totally cyclic* orientations, defined as orientations such that any two vertices in the same connected component of G lie in a directed cycle. It tuns out that if two orientations are equivalent, one is totally cyclic if the other one is. The set of equivalence classes of totally cyclic orientations on G is denoted by $\overline{O}^0(G)$.

In the picture below we have the four orientations on a 2-cycle. The first two are totally cyclic and equivalent. The last two are not totally cyclic.

We shall adopt the convention that the empty orientation on the graph consisting of only one vertex and no edges is totally cyclic. We notice the following facts.

Remark 4.2.1. (a) $\overline{O}^0(G)$ is empty if and only if G contains some bridge.

(b) Assume G connected. Then $|\overline{O}^0(G)| \leq |\Phi_G|$ with equality if and only if $|V| = 1$.

The set of all orientations on the spanning subgraphs of a graph G admits a partial order as follows. Let O_{S_1} and O_{S_2} be two orientations on $G - S_1$ and $G - S_2$ respectively, where $S_i \subset E(G)$ for $i = 1, 2$. We set $O_{S_1} \leq O_{S_2}$ if $G - S_1 \subset G - S_2$ and if the restriction of O_{S_2} to $G - S_1$ equals O_{S_1}.

This definition is compatible with the equivalence relation defined above, and hence the set of all equivalence classes of totally cyclic orientations on G is a poset, which we shall denote as follows

$$\overline{\mathcal{OP}}^0(G) := \bigsqcup_{S \subset E(G)} \overline{O}^0(G - S).$$

Finally, we are ready to exhibit a graded stratification of $\overline{P}_X^{g-1}$ governed by totally cyclic orientations, by rephrasing some results in Caporaso and Viviani [2010].

Proposition 4.2.2. *Let X be a stable curve of genus g and G its dual graph. Then the following is a graded stratification*

$$\overline{P}_X^{g-1} = \bigsqcup_{\overline{O_S}\in\overline{\mathcal{OP}}^0(G)} P_X^{O_S}, \tag{11}$$

and we have natural isomorphisms for every $\overline{O_S} \in \overline{\mathcal{OP}}^0(G)$

$$P_X^{O_S} \cong \operatorname{Pic}^{\underline{d}^{O_S}}(X_S^\nu). \tag{12}$$

The isomorphisms (12) exhibit the recursive behaviour described in Remark 2.3.3. Indeed, $\operatorname{Pic}^{\underline{d}^{O_S}}(X_S^\nu) \cong J_{X_S^\nu}$ and $\underline{d}^{O_S}$ is the multidegree associated to a totally cyclic orientation on $G - S$. Hence the boundary of the compactified degree-$(g-1)$ Jacobian of X is stratified by Jacobians of degree $(g(X')-1)$ of partial normalizations, X', of X.

Remark 4.2.3. If X is reducible and not of compact type then $\overline{P}_X^{g-1}$ is not of Néron type, by Remark 4.2.1(b).

A tropical counterpart of the stratification (11) is not known to us.

The compactified Jacobians in degree $g-1$ have been proved especially useful in various situations and in connection with the Theta divisor, whose definition extends to these compactified Jacobians. Among the applications, we recall that the pair given by this degree-$(g-1)$ Jacobian and its Theta divisor, $(\overline{P}_X^{g-1}, \overline{\Theta}_X)$, is endowed with a natural group action of J_X and, as such, forms a so-called *principally polarized stable semi-abelic pair*. These pairs appear as boundary points in the compactification, $\overline{A}_g^{\text{mod}}$, of the moduli space of principally polarized abelian varieties constructed in Alexeev [2002]. Moreover, by Alexeev [2004], they form the image of the compactified Torelli map and we have

Proposition 4.2.4. *The extended Torelli morphism $\overline{\tau}: \overline{M}_g \to \overline{A}_g^{mod}$ mapping a curve X to $(\overline{P}_X^{g-1}, \overline{\Theta}_X)$ is the moduli map associated to the family $\psi_{g-1}: \overline{P}_g^{g-1} \to \overline{M}_g$.*

The combinatorial structure of $\overline{P}_X^{g-1}$ described in Proposition 4.2.2 is heavily used in Caporaso and Viviani [2010] to study the fibers of $\overline{\tau}$.

4.3 Jacobians in degree g: Néron compactified Jacobians. We now adapt the considerations at the beginning of the previous sections to the case $d = g$. We have $g = \sum_{v\in V}(h(v)-1)+|E|+1$ and, as before, we want to express the term $|E|+1$ in a combinatorially meaningful way. Modifying what we did earlier, we now consider orientations having one bioriented edge. More precisely, a *1-orientation* on a graph G is the datum of a *bioriented* edge, e, and of an orientation on $G - e$. For any 1-orientation we have, of course, $|\underline{d}^O| = g(G)$. Just as in the previous subsection we restricted to totally cyclic orientations, now we need to restrict to certain special 1-orientations, namely "rooted" orientations. A 1-orientation with bioriented edge e is said to be *rooted* (or *e-rooted*) if for every vertex v there exists a directed path from e to v.

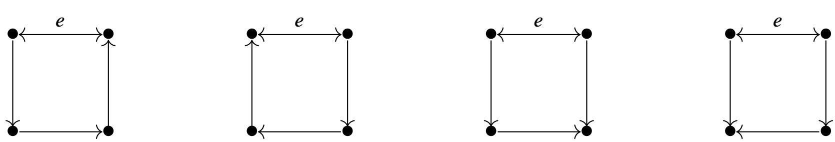
As before, two 1-orientations O and O' such that $\underline{d}^O = \underline{d}^{O'}$ are defined to be equivalent, moreover one is rooted if the other one is. We denote by $\overline{O}^1(G)$ the set of equivalence classes of rooted 1-orientations on G.

We decree the empty orientation on the graph consisting of only vertex and no edges to be rooted. Similarly to Remark 4.2.1 we have:

Remark 4.3.1. (a) $\overline{O}^1(G)$ is not empty if and only if G is connected.

(b) $|\overline{O}^1(G)| = |\Phi_G|$.

In the picture below we have all e-rooted orientations on a 4-cycle with fixed bioriented edge e. They correspond to the four elements in $\overline{O}^1(G)$.



In Theorem 2.3.1 we exhibited a stratification indexed by the poset of connected spanning subgraphs, $\mathcal{C}(G)$. The strata of that stratification are not connected, we shall now exhibit a finer stratification with connected strata.

Recall that rooted 1-orientations exist only on connected graphs. The poset of equivalence classes of rooted 1-orientations on all spanning subgraphs of G is written as follows

$$\overline{\mathcal{OP}}^1(G) := \bigsqcup_{S \in \mathcal{C}(G)} \overline{O}^1(G - S), \tag{13}$$

with the partial order defined in the previous section.

The following result of Christ [2017] states that $\overline{P}^g_X$ admits a recursive graded stratification governed by rooted orientations.

Theorem 4.3.2 (Christ)**.** *Let X be a stable curve of genus g and G its dual graph. Then $\overline{P}^g_X$ admits the following graded stratification*

$$\overline{P}^g_X = \bigsqcup_{\overline{O_S}\in\overline{\mathcal{OP}}^1(G)} P_X^{O_S}, \tag{14}$$

and we have natural isomorphisms for every $\overline{O_S} \in \overline{\mathcal{OP}}^1(G)$

$$P_X^{O_S} \cong \mathrm{Pic}^{\underline{d}^{O_S}}(X_S^\nu). \tag{15}$$

Remark 4.3.3. Comparing with the stratification of Theorem 2.3.1 we have, using (13), that (14) is a refinement of (7), with connected strata.

In this case we do have a tropical counterpart. First, consistently with the dual stratifications (9) and (10), a tropical counterpart of X is a tropical curve Γ whose underlying graph is the dual graph of X.

We now assume $\Gamma = (G, \ell)$ is compact (i.e. not extended). The tropical curve Γ has a Picard group $\mathrm{Pic}(\Gamma) = \sqcup \mathrm{Pic}^d(\Gamma)$, and each connected component, $\mathrm{Pic}^d(\Gamma)$, is isomorphic to the same $b_1(G)$-dimensional real torus; see Mikhalkin and Zharkov [2008]. In An, Baker, Kuperberg, and Shokrieh [2014], the authors show that $\mathrm{Pic}^g(\Gamma)$ has an interesting polyhedral decomposition indexed by "break divisors" on G. The connection between break divisors and rooted 1-orientations is established, as a consequence of the results in An, Baker, Kuperberg, and Shokrieh [ibid.], in Christ [2017], where the following result is obtained.

Theorem 4.3.4. *Let $\Gamma = (G, \ell)$ be a compact tropical curve of genus g. Then $\mathrm{Pic}^g(\Gamma)$ admits the following graded stratification*

$$\mathrm{Pic}^g(\Gamma) = \bigsqcup_{\overline{O_S}\in\overline{\mathcal{OP}}^1(G)^*} \Sigma_\Gamma^{O_S}. \tag{16}$$

The stratification (16) is a non-trivial rephrasing of the polyhedral decomposition established in An, Baker, Kuperberg, and Shokrieh [2014]. Such a rephrasing is needed to establish the connection with the stratification (14). From An, Baker, Kuperberg, and Shokrieh [ibid.] it follows that the strata $\Sigma_\Gamma^{O_S}$ are the interiors of the faces of a polyhedral decomposition for $\mathrm{Pic}^g(\Gamma)$.

Presently, we do not know whether the duality between the stratifications (16) and (14) can be given an interpretation in terms of tropicalization and analytification, similarly to the cases described in subsections 3.2 and 3.3.

This problem is related to a result of Baker and Rabinoff [2015], which we will state in our notation. Using Theorem 3.2.1, let $\mathfrak{X}_K$ be a smooth curve, $J_{\mathfrak{X}_K}$ its Jacobian, and let $\Gamma = \mathrm{trop}([\mathfrak{X}_K])$, where $[\mathfrak{X}_K]$ is the point of $\overline{M}_g^{\mathrm{an}}$ corresponding to $\mathfrak{X}_K$. Hence Γ is a compact tropical curve of genus g (compactness follows from $\mathfrak{X}_K$ being smooth). With this set-up, Baker and Rabinoff [ibid., Thm. 1.3] yields

Theorem 4.3.5 (Baker-Rabinoff)**.** $\mathrm{Pic}^g(\Gamma) \cong \Sigma(J^{\mathrm{an}}_{\mathfrak{X}_K})$.

With this result in mind, a natural approach to the problem mentioned above would be to study the relation between $\Sigma(J^{\mathrm{an}}_{\mathfrak{X}_K})$ and $\overline{P}^g_X$.

Finally, consider the universal compactified Jacobian. Results from Christ [2017] indicate that an analogue of Theorem 4.3.2 should hold uniformly over $\overline{\mathfrak{M}}_g$, so that the universal compactified Jacobian $\overline{\mathfrak{P}}^g_g$ can be given a graded stratification compatible with that of $\overline{\mathfrak{M}}_g$. We expect the same to hold for the universal compactified Jacobian in degree $g-1$, with Proposition 4.2.2 as starting point.

References

Dan Abramovich, Lucia Caporaso, and Sam Payne (2015). "The tropicalization of the moduli space of curves". *Ann. Sci. Éc. Norm. Supér. (4)* 48.4, pp. 765–809. MR: 3377065 (cit. on pp. 659–661).

Valery Alexeev (2002). "Complete moduli in the presence of semiabelian group action". *Ann. of Math. (2)* 155.3, pp. 611–708 (cit. on p. 666).

– (2004). "Compactified Jacobians and Torelli map". *Publ. Res. Inst. Math. Sci.* 40.4, pp. 1241–1265. MR: 2105707 (cit. on p. 666).

Yang An, Matthew Baker, Greg Kuperberg, and Farbod Shokrieh (2014). "Canonical representatives for divisor classes on tropical curves and the matrix-tree theorem". *Forum Math. Sigma* 2, e24, 25. MR: 3264262 (cit. on p. 668).

Matthew Baker, Sam Payne, and Joseph Rabinoff (2013). "On the structure of non-Archimedean analytic curves". In: *Tropical and non-Archimedean geometry*. Vol. 605. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 93–121. MR: 3204269 (cit. on p. 661).

– (2016). "Nonarchimedean geometry, tropicalization, and metrics on curves". *Algebr. Geom.* 3.1, pp. 63–105. MR: 3455421 (cit. on p. 661).

Matthew Baker and Joseph Rabinoff (2015). "The skeleton of the Jacobian, the Jacobian of the skeleton, and lifting meromorphic functions from tropical to algebraic curves". *Int. Math. Res. Not. IMRN* 16, pp. 7436–7472. MR: 3428970 (cit. on pp. 668, 669).

Arnaud Beauville (1977). "Prym varieties and the Schottky problem". *Invent. Math.* 41.2, pp. 149–196. MR: 0572974 (cit. on p. 664).

Vladimir Berkovich (1990). *Spectral theory and analytic geometry over non-Archimedean fields*. Vol. 33. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, pp. x+169. MR: 1070709 (cit. on p. 660).

Silvia Brannetti, Margarida Melo, and Filippo Viviani (2011). "On the tropical Torelli map". *Adv. Math.* 226.3, pp. 2546–2586. MR: 2739784 (cit. on pp. 659, 660).

Lucia Caporaso (1994). "A compactification of the universal Picard variety over the moduli space of stable curves". *J. Amer. Math. Soc.* 7.3, pp. 589–660. MR: 1254134 (cit. on p. 657).

– (2008). "Néron models and compactified Picard schemes over the moduli stack of stable curves". *Amer. J. Math.* 130.1, pp. 1–47. MR: 2382140 (cit. on p. 657).

– (2012a). "Algebraic and tropical curves: comparing their moduli spaces". In: *Handbook of Moduli, Volume I*. Vol. XXIV. Advanced Lectures in Mathematics. International Press, Boston, pp. 119–160 (cit. on pp. 659, 660).

– (2012b). "Compactified Jacobians of Néron type". *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* 23.2, pp. 213–227. MR: 2924900 (cit. on p. 657).

Lucia Caporaso and Filippo Viviani (2010). "Torelli theorem for graphs and tropical curves". *Duke Math. J.* 153.1, pp. 129–171. MR: 2641941 (cit. on pp. 665, 666).

Renzo Cavalieri, Hannah Markwig, and Dhruv Ranganathan (2016). "Tropicalizing the space of admissible covers". *Math. Ann.* 364.3-4, pp. 1275–1313. MR: 3466867 (cit. on pp. 661, 662).

Karl Christ (2017). "Orientations, break Divisors and compactified Jacobians". *PhD thesis in Mathematics, Roma Tre University*. arXiv: 1604.03176 (cit. on pp. 667–669).

Pierre Deligne and David Mumford (1969). "The irreducibility of the space of curves of given genus". *Inst. Hautes Études Sci. Publ. Math.* 36, pp. 75–109. MR: 0262240 (cit. on p. 659).

Eduardo Esteves (2001). "Compactifying the relative Jacobian over families of reduced curves". *Trans. Amer. Math. Soc.* 353.8, 3045–3095 (electronic). MR: 1828599 (cit. on p. 657).

David Gieseker (1982). *Lectures on moduli of curves*. Vol. 69. Tata Institute of Fundamental Research Lectures on Mathematics and Physics. Published for the Tata Institute of Fundamental Research, Bombay; Springer-Verlag, Berlin-New York, pp. iii+99. MR: 691308 (cit. on pp. 659, 663).

Walter Gubler, Joseph Rabinoff, and Annette Werner (2016). "Skeletons and tropicalizations". *Adv. Math.* 294, pp. 150–215. MR: 3479562 (cit. on p. 661).

Finn F. Knudsen (1983a). "The projectivity of the moduli space of stable curves. II. The stacks $M_{g,n}$". *Math. Scand.* 52.2, pp. 161–199. MR: 702953 (cit. on p. 659).

– (1983b). "The projectivity of the moduli space of stable curves. III. The line bundles on $M_{g,n}$, and a proof of the projectivity of $\overline{M}_{g,n}$ in characteristic 0". *Math. Scand.* 52.2, pp. 200–212. MR: 702954 (cit. on p. 659).

Diane Maclagan and Bernd Sturmfels (2015). *Introduction to tropical geometry*. Vol. 161. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, pp. xii+363. MR: 3287221 (cit. on p. 660).

Margarida Melo, Antonio Rapagnetta, and Filippo Viviani (2017). "Fine compactified Jacobians of reduced curves". *Trans. Amer. Math. Soc.* 369.8, pp. 5341–5402. MR: 3646765 (cit. on p. 657).

Margarida Melo and Filippo Viviani (2012). "Fine compactified Jacobians". *Math. Nachr.* 285.8-9, pp. 997–1031. MR: 2928396 (cit. on p. 657).

Grigory Mikhalkin (2007). "Moduli spaces of rational tropical curves". In: *Proceedings of Gökova Geometry-Topology Conference 2006*. Gökova Geometry/Topology Conference (GGT), Gökova, pp. 39–51. MR: 2404949 (cit. on p. 660).

Grigory Mikhalkin and Ilia Zharkov (2008). "Tropical curves, their Jacobians and theta functions". In: *Curves and abelian varieties*. Vol. 465. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 203–230. MR: 2457739 (cit. on pp. 660, 668).

André Néron (1964). "Modèles minimaux des variétés abéliennes sur les corps locaux et globaux". *Inst. Hautes Études Sci. Publ.Math. No.* 21, p. 128. MR: 0179172 (cit. on p. 655).

Tadao Oda and C. S. Seshadri (1979). "Compactifications of the generalized Jacobian variety". *Trans. Amer. Math. Soc.* 253, pp. 1–90. MR: 536936 (cit. on pp. 656, 657, 663).

S. Payne (2009). "Analytification is the limit of all tropicalizations". *Math. Res. Lett.* 16.3, pp. 543–556. MR: 2511632 (cit. on p. 660).

Michel Raynaud (1970). "Spécialisation du foncteur de Picard". *Inst. Hautes Études Sci. Publ. Math.* 38, pp. 27–76. MR: 0282993 (cit. on pp. 655, 656).

Amaury Thuillier (2007). "Géométrie toroïdale et géométrie analytique non archimédienne. Application au type d'homotopie de certains schémas formels". *Manuscripta Math.* 123.4, pp. 381–451. MR: 2320738 (cit. on p. 660).

Ilya Tyomkin (2012). "Tropical geometry and correspondence theorems via toric stacks". *Math. Ann.* 353.3, pp. 945–995 (cit. on p. 661).

Filippo Viviani (2013). "Tropicalizing vs. compactifying the Torelli morphism". In: *Tropical and non-Archimedean geometry*. Vol. 605. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 181–210. MR: 3204272 (cit. on p. 661).

Received 2017-11-30.

Lucia Caporaso
Dipartimento di Matematica e Fisica
Università Roma Tre
Largo San Leonardo Murialdo
I-00146 Roma
Italy
caporaso@mat.uniroma3.it

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (671–688)

ON EXPLICIT ASPECT OF PLURICANONICAL MAPS OF PROJECTIVE VARIETIES

Jungkai A. Chen (陳榮凱) and Meng Chen (陈猛)

Abstract

In this survey article, we introduce the development of birational geometry associated to pluricanonical maps. Especially, we explain various aspects of explicit studies of threefolds including the key idea of theory of baskets and other applications.

1 Introduction

In the realm of birational geometry, divisors are perhaps the most important objects. Classically they were used to keep track of the property of zeros and poles and naturally developed into a convenient tool to study functions with preassigned conditions. Thus one may study linear systems associated to various interesting divisors. Among all divisors, the canonical divisor K, together with m-th canonical divisor mK for any $m \in \mathbb{Z}$, plays the central role. The behavior of pluricanonical maps φ_m or pluricanonical systems $|mK|$ is intensively studied in the minimal model program (in short, MMP) as well as other "canonical" classification problems. In fact, many very important concepts in algebraic geometry such as Kodaira dimension, Iitaka fibration, canonical volume, extremal contractions in minimal model program and so on, are defined on the basis of specific properties of the canonical divisor K.

More explicitly let V be any nonsingular projective variety of dimension n. For any $m \in \mathbb{Z}$, denote by $\varphi_{m,V}$ the m-canonical map of V. The following three types of problems are in the core of birational geometry:

Question 1.1. For any integer $n \geq 3$, find a practical integer r_n so that, for all nonsingular projective n-folds of general type, φ_m is birational onto its image for all $m \geq r_n$.

The first author was partially supported by NCTS and Ministry of Science and Technology of Taiwan. The second author was supported by National Natural Science Foundation of China (#11571076, #11231003, #11421061) and Program of Shanghai Subject Chief Scientist (#16XD1400400).

MSC2010: primary 14J30; secondary 14E99.

Question 1.2. For any integers $n \geq 3$ and $n > \kappa \geq 0$, find integers $M_{n,\kappa}$ and $d_{n,\kappa}$ such that, for all nonsingular projective n-folds with Kodaira dimension κ, the m-th canonical map φ_m defines an Iitaka fibration for all $m \geq M_{n,\kappa}$ and divisible by $d_{n,\kappa}$.

Question 1.3. For any integer $n \geq 3$, find an integer m_n so that, for all canonical (terminal) weak $\mathbb{Q}$-Fano n-folds (i.e. $-K$ being $\mathbb{Q}$-Cartier, nef and big), φ_{-m} is birational onto its image for all $m \geq m_n$.

First we recall some results on the quest of existence for r_n, $M_{n,\kappa}$ and m_n.

- For varieties V of general type, the remarkable theorem, proved separately by Hacon and McKernan [2006], Takayama [2006] and Tsuji [2006], asserts that there is a constant $\tilde{r}_n$ depending only on the dimension n ($n > 2$) such that $\varphi_{m,V}$ is birational for all $m \geq \tilde{r}_n$.

- For varieties of intermediate Kodaira dimension, say $0 < \kappa(V) < n$, the effective Iitaka fibration conjecture predicts that there exists a constant $\tilde{c}_n$ depending only on n such that $\varphi_{m,V}$ defines an Iitaka fibration for all $m \geq \tilde{c}_n$ and divisible. The canonical bundle formula of Fujino and Mori [2000] serves as the fundamental tool in this situation. The up-to-date result, due to Birkar and Zhang [2016] says that there exists a uniform number $M(n, b_F, \beta_{\tilde{F}})$ so that φ_m gives an Iitaka fibration for all $m \geq M(n, b_F, \beta_{\tilde{F}})$ and divisible. The numbers b_F and $\beta_{\tilde{F}}$ are defined as follows. Let F be the general fiber of Iitaka fibration. The number b_F, called the *index of fiber*, is the smallest positive integer so that $|bK_F| \neq \varnothing$. One has a covering $\tilde{F} \to F$ by $|mK_F|$. Then $\beta_{\tilde{F}}$, called the *middle Betti number*, is defined as the $(n-\kappa)$-th Betti number of the $n-\kappa$ dimensional variety $\tilde{F}$. One may refer to Viehweg and Zhang [2009], G. Todorov and Xu [2009], Pacienza [2009], X. Jiang [2013], Di Cerbo [2014] and Birkar and Zhang [2016] for more details along this direction.

- For canonical (resp. terminal) weak $\mathbb{Q}$-Fano threefolds, the boundedness was proved by Kawamata [1992] under the condition that the Picard number $\rho = 1$ and by Kollár, Miyaoka, Mori, and Takagi [2000] for the general case with $\rho > 1$. Recent breakthrough of Birkar [2016] asserts that even for $n \geq 4$ there is a constant $\tilde{m}_n$ depending only on n such that $\varphi_{-m,V}$ is birational for all $m \geq \tilde{m}_n$.

It is interesting to study the explicit aspect of pluricanonical maps of projective varieties in high dimensions. Some recent advances show that r_3, m_3, and $M_{3,\kappa}$ have realistic bounds which are very close to being optimal. The purpose of this survey article is to introduce and to sketch some of the key ideas and techniques developed from those explicit studies of 3-folds. We expect that such detailed and explicit studies of 3-folds will pave a solid path toward the understanding of higher dimensional birational geometry.

Throughout, all varieties are considered over an algebraically closed field k of characteristic zero.

2 Theory of weighted baskets

The understanding of terminal and canonical singularities plays the essential role in the development of three dimensional geometry. The milestone work of the existence of flips in dimension three, due to Mori [1988], built on the classification of terminal singularities and extremal neighborhoods. Moreover, Reid showed that each terminal singularities can be deformed into cyclic quotient singularities. Hence, the collection of deformed quotient singularities carries ample information of singularities. This leads to the notion of *baskets of terminal orbifold points*. For simplicity, a terminal orbifold point of type $\frac{1}{r}(1,-1,b)$ will be denoted as (b,r) with $b \le r/2$. A basket, which is a collection of terminal orbifold points, is written as $\mathcal{B} = \{n_i \times (b_i, r_i)\}$ where n_i denotes the multiplicities.

The subsequent result of singular Riemann-Roch formula (see Reid [1987]) can be derived by computing the contribution of basket of singularities, say

$$\chi(\mathcal{O}_X(D)) = \chi(\mathcal{O}_X) + \frac{1}{12}D(D-K_X)(2D-K_X) + \frac{1}{12}(D.c_2(X)) + \sum_{P\in B(X)} \left(-i_P \cdot \frac{r_P^2-1}{12r_P} + \sum_{j=1}^{i_P-1} \frac{\overline{jb_P}(r_P-\overline{jb_P})}{2r_P} \right),$$

where $c_2(X)$ is defined in the sense of intersection theory, $B(X) = \{(b_P, r_P)\}$ is the basket data of X and i_P is the local index of D such that $\mathcal{O}_X(D) \cong \mathcal{O}_X(i_P K_X)$ near P.

By applying the singular Riemann-Roch formula to $D = K_X$, then one gets

$$(K_X.c_2(X)) = -24\chi(\mathcal{O}_X) + \sum_{P\in B_X} \left(r_P - \frac{1}{r_P} \right).$$

This leads to various results. For example, Kawamata and Morrison's result on the global index of 3-folds with $K_X \equiv 0$ was then derived (see Section 5).

By taking $D = mK_X$ and replacing $(K_X.c_2(X))$ with $\chi(\mathcal{O}_X)$ and the contribution of singularities, we get the following plurigenus formula Reid [ibid.]:

$$\chi_m = \frac{1}{12}m(m-1)(2m-1)K^3 + (1-2m)\chi + l(m), \tag{2-1}$$

where $\chi = \chi(\mathcal{O}_X)$, $K^3 = K_X^3$, $\chi_m = \chi(\mathcal{O}_X(mK_X))$ and

$$l(m) = \sum_{P\in B_X} \sum_{j=1}^{m-1} \frac{\overline{jb_P}(r_P-\overline{jb_P})}{2r_P}. \tag{2-2}$$

It is clear from the Riemann-Roch formula that the triple (B_X, χ_2, χ) determine χ_m for all $m \geq 3$. We call the triple $\mathbb{B} = \{B, \chi_2, \chi\}$ a *weighted basket*, where B is a basket of orbifold points, χ_2 is a non-negative integer and χ is an integer. For any $m \geq 3$, χ_m can be inductively and formally defined by means of (2-1). Note that the rational number K^3, which is also uniquely determined by $\mathbb{B}$, is called the volume of $\mathbb{B}$.

Given a basket

$$B = \{(b_1, r_1), (b_2, r_2), \dots, (b_k, r_k)\},$$

we call the basket

$$B' := \{(b_1 + b_2, r_1 + r_2), (b_3, r_3), \dots, (b_k, r_k)\}$$

a packing of B, written as $B \succcurlyeq B'$. If $b_1 r_2 - b_2 r_1 = 1$, then we call $B \succcurlyeq B'$ a *prime packing*.

Then we introduced the "canonical sequence of prime unpackings of a basket" in J. A. Chen and M. Chen [2010b]:

$$B^{(0)}(B) \succcurlyeq B^{(5)}(B) \succcurlyeq \dots \succcurlyeq B^{(n)}(B) \succcurlyeq \dots \succcurlyeq B,$$

so that $B^{(n)}$ consists of orbifold points (b_i, r_i) with either $r_i \leq n$ or $b_i = 1$. The basket $B^{(0)}$, called *the initial basket*, consists of orbifold points of the form $(1, r_i)$. Take an orbifold point (b, r) for example. Let $q = \lfloor \frac{r}{b} \rfloor$. Then $\frac{1}{q} \geq \frac{b}{r} \geq \frac{1}{q+1}$. Indeed, let $\beta := r - qb$, then the initial basket of (b, r) is $\{(b - \beta) \times (1, q), \beta \times (1, q+1)\}$.

The packing of baskets naturally induces the packing of weighted baskets, namely we define

$$\{B, \chi_2, \chi\} \succcurlyeq \{B', \chi_2, \chi\}$$

if $B \succcurlyeq B'$. Furthermore, any given weighted basket $\mathbb{B}$ has the corresponding "canonical sequence":

$$\mathbb{B}^{(0)} \succcurlyeq \mathbb{B}^{(5)} \succcurlyeq \dots \succcurlyeq \mathbb{B}^{(n)} \succcurlyeq \dots \succcurlyeq \mathbb{B}.$$

As revealed in J. A. Chen and M. Chen [ibid., Section 3], the intrinsic properties of the canonical sequence provide many new inequalities among the Euler characteristic and other characteristics of a given weighted basket $\mathbb{B}$, of which the most interesting one is:

$$2\chi_5 + 3\chi_6 + \chi_8 + \chi_{10} + \chi_{12} \geq \chi + 10\chi_2 + 4\chi_3 + \chi_7 + \chi_{11} + \chi_{13} + R, \tag{2-3}$$

where R is certain non-negetative combination of all initial baskets with higher indices.

Although the above notions were introduced in a very formal way, it was proved to be quite effective for various geometric problems, which we will discuss in next sections. Also, even though the notion of packings was introduced to study the numerical behavior rather than its geometric meaning at the beginning. The relation appears in divisorial contractions to points.

Example 2.1. Let $P = \frac{1}{9}(2,7,1) \in X \cong \mathbb{C}^3/\mu_9$ be a terminal quotient singularity. Let $Y \to X$ be the weighted blowup with weights $\frac{1}{9}(2,7,1)$. Then the basket of Y is $B_Y = \{(1,2),(3,7)\}$ and the basket of X is $\{(4,9)\}$ which is a packing of B_Y. This is an example of *Kawamata blowup* (cf. Kawamata [1996]).

Example 2.2. Let $P \in X \cong (xy + z^{15} + u^2 = 0) \subset \mathbb{C}^4/\mu_5$ (of type $\frac{1}{5}(3,2,1,5)$) be a $cA/5$ singularity. Let $Y \to X$ be a weighted blowup with weights $\frac{1}{5}(3,7,1,5)$. Then the basket of Y is $\{(3,7),(1,3)\}$ and the basket of X is $\{2 \times (2,5)\}$. They have the same initial baskets $\{2 \times (1,2), 2 \times (1,3)\}$.

3 Pluricanonical maps of threefolds of general type

We consider 3-folds of general type in this section. Since minimal models exist and the problems are birational in nature, we usually work on minimal projective 3-folds of general type with $\mathbb{Q}$-factorial terminal singularities, unless otherwise stated. Denote by r_X the Cartier index of X. Define the *canonical stability index*

$$r_s(X) = \min\{t \mid \varphi_{m,X} \text{ is birational for all } m \geq t\}.$$

Clearly $r_3 = \max\{r_s(X) \mid X \text{ is a minimal 3-fold of general type}\}$.

3.1 The case $r_X = 1$.

When X is smooth and minimal, it is Wilson [1980] who first proved that $r_s(X) \leq 25$. Then this was improved, chronoligically, by Benveniste [1986] ($r_s(X) \leq 9$), Matsuki [1986] ($r_s(X) \leq 7$) and the second author M. Chen [1998] ($r_s(X) \leq 6$).

In fact, the method of Benveniste, Matsuki and the second author can be easily extended to the situation, when X is Gorenstein and minimal, by using a special partial resolution due to Reid [1983] and Miyaoka [1987]. Note also that S. Lee [2000] proved the optimal base point freeness of $|4K|$.

Example 3.1. Let $X = S \times C$ where S is a minimal surface of general type of $(1,2)$-type and C is a complete curve of genus ≥ 2. Then $r_s(X) = 5$ according to Bombieri [1973]. Thus, for Gorenstein minimal 3-folds of general type, the best one can expect is that $r_s(X) \leq 5$ holds.

Question 1.1 in the case of $n = 3$ and $r_X = 1$ was finally solved in 2007 by the authors and De-Qi Zhang:

Theorem 3.2. *(J. A. Chen, M. Chen, and Zhang [2007, Theorem 1.1]) Let X be a minimal projective 3-fold of general type with $r_X = 1$. Then $\varphi_{m,X}$ is a biratioal morphism for every integer $m \geq 5$.*

3.2 Kollár's method.

The work of Kollár on push-forward of dualizing sheaves provide various important applications in the study of higher dimensional geometry. One of the them is to reduce the birationality problem to non-vanishing of plurigenera as in the following theorem:

Theorem 3.3. *(Kollár [1986, Corollary 4.8]) Let V be a nonsingular projective 3-fold of general type with $P_k(V) \geq 2$ for some integer $k > 0$. Then φ_{11k+5} is birational.*

The key contribution of Theorem 3.3 in the process of solving Question 1.1 ($n = 3$) is that it reduces to the problem to find an effective integer k so that $P_k \geq 2$, which is the standard task of Riemann-Roch formula.

Kollár's method of proving Theorem 3.3 is as follows. If one takes a sub-pencil $\Lambda \subset |kK_V|$, modulo a further birational modification if necessary, one gets a surjective morphism $f : V \longrightarrow \Gamma \cong \mathbb{P}^1$. One has the inclusion $\mathcal{O}(1) \hookrightarrow f_*\omega_V^k$ and then, for any $p \geq 5$,

$$f_*\omega_{V/\Gamma}^p \otimes \mathcal{O}(1) \hookrightarrow f_*\omega_V^{(2p+1)k+p}.$$

Since the 5-canonical map of the general fiber is birational and by the semi-positivity of $f_*\omega_{V/\Gamma}^p$, one sees that φ_{11k+5} is birational by simply taking $p = 5$.

Theorem 3.3 was considerably improved by the second author M. Chen [2004a, Theorem 0.1] that, under the same condition as that of Theorem 3.3, φ_{5k+6} is birational. Some further optimal results were proved in M. Chen [2003], M. Chen [2004a], M. Chen [2007], and J. A. Chen and M. Chen [2010a].

3.3 The case of $r_X \geq 2$.

Turning to the general situation that minimal model contains some singularities of index ≥ 2. Suppose $\chi(\mathcal{O}_X) < 0$. Reid's Riemann-Roch formula implies $P_2(X) \geq 4$. Hence the question is solvable by Kollár's theorem. Suppose that $P_k \geq 2$ for some $k \leq 12$, one can apply Kollár's method as well.

It remains to consider $\chi(\mathcal{O}_X) \geq 0$ and $P_k(X) \leq 1$ for all $2 \leq k \leq 12$. The key inequality (2-3) reads:

$$2P_5 + 3P_6 + P_8 + P_{10} + P_{12} \geq \chi(\mathcal{O}_X) + 10P_2 + 4P_3 + P_7 + P_{11} + P_{13}, \tag{3-1}$$

which directly implies that $\chi(\mathcal{O}_X) \leq 8$. This also means that P_{13} is upper bounded by 7. In practice, one may obtain more precise bounds for both $\chi(\mathcal{O}_X)$ and P_{13}. It turns out that the 12-th weighted basket $\mathbb{B}^{12}(X)$ has only finite possibilities and so does $\mathbb{B}(X)$. It is then possible to answer Question 1.1, which was the main work in the authors' papers J. A. Chen and M. Chen [2010b,a, 2015a].

To improve or to reach the possible optimal bound, one needs to study the birational geometry explicitly. Known useful techniques include some effective method to estimate the lower bound of K_X^3 and better birationality criterion of $\varphi_{m,X}$ under the assumption that $P_{m_0} \geq 2$ for some number $m_0 > 0$, which is a kind of improvement to Kollár's method. We briefly describe the technique below. Indeed, if $P_{m_0} \geq 2$, then one has an induced fibration from a sub-pencil of $|m_0 K|$, say $f : X' \longrightarrow \Gamma$, where X' is a smooth model of X and Γ is a smooth complete curve. Denote by $\pi : X' \longrightarrow X$ the birational morphism. Pick a general fiber F of f and denote by $\sigma : F \longrightarrow F_0$ the contraction onto its minimal model. Naturally one has

$$m_0 \pi^*(K_X) \equiv pF + E_{m_0},$$

for some integer $p > 0$ and effective $\mathbb{Q}$-divisor E_{m_0}. The key point is to prove the so-called "canonical restriction inequality" (see M. Chen [2003], M. Chen [2004a], M. Chen [2007], M. Chen and Zhang [2008], M. Chen and Zuo [2008], and J. A. Chen and M. Chen [2010a, 2015a] for its development history) as follows:

$$\pi^*(K_X)|_F \geq \frac{p}{m_0 + p} \sigma^*(K_{F_0}) \tag{3-2}$$

modulo $\mathbb{Q}$-linear equivalence. The inequality (3-2) directly gives an effective lower bound of K_X^3 and it is crucial as well in proving the birationality of $\varphi_{m,X}$.

Here are the main theorems of the authors as the answer to Question 1.1 ($n = 3$):

Theorem 3.4. *(J. A. Chen and M. Chen [2010b,a, 2015a]) Let X be a minimal projective 3-fold of general type. Then*

(1) $K_X^3 \geq \frac{1}{1680}$;

(2) $\varphi_{m,X}$ *is birational for all* $m \geq 61$;

(3) $P_{12} \geq 1$ *and* $P_{24} \geq 2$.

(4) $K_X^3 \geq \frac{1}{420}$ *(optimal) if* $\chi(\mathcal{O}_X) \leq 1$.

Remark 3.5. The statements are birational in nature. More precisely, φ_m are birationally equivalent on different birational models and the canonical volume $\mathrm{Vol}(V)$ is a birational invariant, equals to K_X^3 of a minimal model X. Therefore, the above statements can be also read as: Let V be a nonsingular projective 3-fold of general type. Then $\mathrm{Vol}(V) \geq \frac{1}{1680}$ and $\varphi_{m,V}$ is birational for all $m \geq 61$, etc.

Define the *pluricanonical section index* $\delta(X)$ to be the minimal integer so that $P_\delta \geq 2$. The author proved the following results:

Theorem 3.6. *(J. A. Chen and M. Chen [2010b,a, 2015a]) Let X be a minimal projective 3-fold of general type. Then*

(1) $\delta(X) \leq 18$*;*

(2) $\delta(X) = 18$ *if and only if* $\mathbb{B}(X) = \{B_{2a}, 0, 2\}$*;*

(3) $\delta(X) \neq 16,\ 17$*;*

(4) $\delta(X) = 15$ *if and only if* $\mathbb{B}(X)$ *belongs to one of the types in J. A. Chen and M. Chen [2015a,* Table F–1*];*

(5) $\delta(X) = 14$ *if and only if* $\mathbb{B}(X)$ *belongs to one of the types in J. A. Chen and M. Chen [ibid.,* Table F–2*];*

(6) $\delta(X) = 13$ *if and only if* $\mathbb{B}(X) = \{B_{41}, 0, 2\}$

where

$$B_{2a} = \{4 \times (1,2), (4,9), (2,5), (5,13), 3 \times (1,3), 2 \times (1,4)\} \text{ and}$$
$$B_{41} = \{5 \times (1,2), (4,9), 2 \times (3,8), (1,3), 2 \times (2,7)\}.$$

Example 3.7. Consider Fletcher's example, which is a general weighted hypersurface X_{46} of weighted degree 46 in weighted projective space $\mathbb{P}(4,5,6,7,23)$ (cf. Iano-Fletcher [2000]). Note that φ_{26} is not birational, $\chi(\mathcal{O}_X) = 1$ and $K_X^3 = \frac{1}{420}$. This provides an example of 3-fold for Theorem 3.4(4).

Recently the second author M. Chen [2016] showed $r_3 \leq 57$ on the basis of above classifications. Therefore, $27 \leq r_3 \leq 57$.

For 3-folds with $\delta = 1$, the second author proved the following optimal results:

Theorem 3.8. *(M. Chen [2003, 2007]) Let X be a minimal projective 3-fold of general type with $p_g(X) \geq 2$. Then*

(1) $K_X^3 \geq \frac{1}{3}$*;*

(2) $\varphi_{8,X}$ *is birational onto its image.*

3.4 On irregular 3-folds of general type. Let X be a minimal projective 3-fold of general type with $q(X) > 0$. One may consider the Albanese map of X. A pioneer work on this topic was due to J. A. Chen and Hacon [2002], who developed the Fourier-Mukai theory to study irregular varieties and proved the following theorem:

Theorem 3.9. *(J. A. Chen and Hacon [2002, 2007]) Let X be a minimal irregular 3-fold of general type. Then*

(1) $|mK_X + P|$ *gives a birational map for all* $m \geq 7$ *(resp.* $m \geq 5$*) and for all (resp. general)* $P \in \mathrm{Pic}^0(X)$.

(2) when $\chi(\omega_X) > 0$, $|mK_X + P|$ *gives a birational map for all* $m \geq 5$ *and for all* $P \in \mathrm{Pic}^0(X)$.

Theorem 3.9(ii) is clearly optimal. For Theorem 3.9(i), the authors and Jiang proved the following theorem:

Theorem 3.10. *(J. A. Chen, M. Chen, and Zhang [2007]) Let* X *be a minimal irregular 3-fold of general type. Then* $\varphi_{6,X}$ *is birational.*

Besides, the authors gave the following effective lower bound for K_X^3:

Theorem 3.11. *(J. A. Chen and M. Chen [2008b, Corollary 1.2]) Let* X *be a minimal irregular 3-fold of general type. Then* $K_X^3 \geq \frac{1}{22}$.

Question 3.12. Are the results in Theorem 3.10 and Theorem 3.11 optimal?

4 The anti-canonical geometry of $\mathbb{Q}$-Fano 3-folds

A normal projective 3-fold X is called a *weak* $\mathbb{Q}$*-Fano 3-fold* (resp. $\mathbb{Q}$*-Fano 3-fold*) if the anti-canonical divisor $-K_X$ is nef and big (resp. ample). A *canonical* (resp. *terminal)* weak $\mathbb{Q}$-Fano 3-fold is a weak $\mathbb{Q}$-Fano 3-fold with at worst canonical (resp. terminal) singularities.

Weak $\mathbb{Q}$-Fano varieties form a fundamental class in minimal model program and various aspects of birational geometry. Given a canonical weak $\mathbb{Q}$-Fano 3-fold X, the *m-th anti-canonical map* $\varphi_{-m,X}$ (or simply φ_{-m}) is the rational map defined by the linear system $|-mK_X|$. It is worthwhile to mention that the behavior of $\varphi_{-m,X}$ in not necessarily birationally invariant, which makes Question 1.3 much harder to work on.

We may always study on a terminal weak $\mathbb{Q}$-Fano 3-fold. Take the weighted basket

$$\mathbb{B}(X) = \{B_X, P_{-1}, \chi(\mathcal{O}_X)\}.$$

By the duality and the vanishing of higher cohomology, we have $\chi_m = -P_{-(m-1)}$ for all $m \geq 2$. Hence the basket theory introduced in Section 2 has a parallel version in Fano case. In fact, the basket theory works very well in classifying weak $\mathbb{Q}$-Fano 3-folds with small invariants.

4.1 Lower bound of the anti-canonical volume.

In 2008, the authors applied the basket theory to prove the following theorem:

Theorem 4.1. *(J. A. Chen and M. Chen [2008a, Theorem 1.1]) Let X be a terminal (or canonical) weak $\mathbb{Q}$-Fano 3-fold. Then*

(1) $P_{-4} > 0$ with possibly one exception of a basket of singularities;

(2) $P_{-6} > 0$ and $P_{-8} > 1$;

(3) $-K_X^3 \geq \frac{1}{330}$. Furthermore $-K_X^3 = -\frac{1}{330}$ if and only if the basket of singularities is $\{(1,2),(2,5),(1,3),(2,11)\}$.

Theorem 4.1.(3) is optimal according to the example of general hypersurface $X_{66} \subset \mathbb{P}(1,5,6,22,33)$ (cf. Iano-Fletcher [2000]).

4.2 The anti-pluricanonical birationality.

The second author started to study the constant for anti-pluricanonical birationality in M. Chen [2011] in 2011. A practical upper bound for m_3 for $\mathbb{Q}$-Fano 3-folds with Picard number one was proved by the second author and C. Jiang in 2016:

Theorem 4.2 (M. Chen and C. Jiang [2016, Theorem 1.6])**.** *Let X be a terminal $\mathbb{Q}$-Fano 3-fold of Picard number one. Then $\varphi_{-m,X}$ is birational for all $m \geq 39$.*

Theorem 4.3 (M. Chen and C. Jiang [ibid., Theorem 1.8, Remark 1.9])**.** *Let X be a canonical weak $\mathbb{Q}$-Fano 3-fold. Then $\varphi_{-m,X}$ is birational for all $m \geq 97$.*

The result in Theorem 4.2 is very close to be optimal according to Fletcher's example Iano-Fletcher [2000]. The numerical bound in Theorem 4.3, however, might be far from optimal. Recently the second author and Jiang had an improvement on this problem:

Theorem 4.4 (M. Chen and C. Jiang [2017, Theorem 1.9])**.** *Let V be a canonical weak $\mathbb{Q}$-Fano 3-fold. Then there exists a terminal weak $\mathbb{Q}$-Fano 3-fold X birational to V such that*

1. *$\dim \overline{\varphi_{-m}(X)} > 1$ for all $m \geq 37$;*
2. *$\varphi_{-m,X}$ is birational for all $m \geq 52$.*

Theorem 4.5 (M. Chen and C. Jiang [ibid., Theorem 1.10])**.** *Let V be a canonical weak $\mathbb{Q}$-Fano 3-fold. Then, for any K-Mori fiber space Y of V,*

1. *$\dim \overline{\varphi_{-m}(Y)} > 1$ for all $m \geq 37$;*
2. *$\varphi_{-m,Y}$ is birational for all $m \geq 52$.*

One notes that an intensive classification using the basket theory developed by the authors was done in proving Theorem 4.4 and Theorem 4.5.

It is a very interesting question to ask what the optimal value of m_3 is, which is crucial in studying the anti-canonical geometry of weak $\mathbb{Q}$-Fano 3-folds. One only knows $33 \le m_3 \le 97$ so far.

5 Threefolds with Kodaira dimension $0 \le \kappa < 3$

For varieties of Kodaira dimension 0, the question is then to find a uniform bound $M_{n,0}$ such that $|mK| \ne \varnothing$ for all m divisible by $M_{n,0}$ and for all n-dimensional varieties with $\kappa = 0$. It is well-known that $M_{2,0} = 12$.

For threefolds with Kodaira dimension 0, Kawamata proved that $0 \le \chi(\mathcal{O}_X) \le 4$. Comparing $\chi(\mathcal{O}_X)$ with $(K_X \cdot c_2)$, one knows the indices of singularities in its minimal model and hence it follows that a uniform bound $M_{3,0}$ exists (cf. Kawamata [1986]). By careful classification of possible singularities, Morrison shows that $|mK_X| \ne \varnothing$ if m is divisible by the Beauville's number (cf. Morrison [1986])

$$\mathrm{lcm}\{m|\phi(m) \le 20\} = 2^5 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19.$$

Note that 20 is chosen as $b_3(A)$ for any abelian threefold A. Indeed, by Oguiso's examples (cf. Oguiso [1993]), the minimal universal number $M_{3,0}$ is the Beauville's number.

For threefolds with $\kappa = 2$, Ringer shows that φ_m is birational to the Iitaka fibration as long as $m \ge 48$ and divisible by 12 (cf. Ringler [2007]). The explicit effective result for threefolds with $\kappa = 1$ was obtained by Hsin-Ku Chen very recently. He show that φ_m is birational to the Iitaka fibration as long as $m \ge 96$ and divisible by 12 (cf. H.-K. Chen [2017]) and hence $M_{3,1} \le 96$ and $d_{3,1} = 12$. There exists an example which shows that $M_{3,1} \ge 42$.

6 Further applications of the theory of baskets and the singular Riemann-Roch

6.1 Weighted complete intersections.

A weighted projective space is a natural generalization of projective spaces. Together with complete intersection inside them, weighted projective spaces provide ample examples. In Iano-Fletcher [2000], Fletcher gave a very detailed account of getting well-formed weighted complete intersection threefolds, including the classification of:

- canonically embedded codimension 1 and 2 weighted 3-folds with total weights < 100;

- weighted complete intersection $\mathbb{Q}$-Fano 3-folds of codimensions 1, 2 with total weights < 100.

Let $X = X_{d_1,\ldots,d_c} \subset \mathbb{P}(a_0, \ldots, a_n)$ be a weighted complete intersection. Let $\delta_j := d_j - a_{j+\dim X}$. By modifying Fletcher's proof, it's not difficult to verify that quasi-smoothness implies that $\delta_j \geq a_{j-1}$. Then the following Theorem follows.

Theorem 6.1. *(J.-J. Chen, J. A. Chen, and M. Chen [2011, Theorem 1.3]) There is no quasi-smooth (not an intersection of a linear cone with another subvariety) weighted complete intersection $X_{d_1,\ldots,d_c} \subset \mathbb{P}(a_0, \ldots, a_n)$ of codimension c greater than* $\dim X + 1 + \alpha$*, where $\alpha = \sum_j d_j - \sum_i a_i$.*

In order to classify 3-fold weighted complete intersections with at worst terminal singularities, let $\mu_i := \#\{a_j | a_j = i\}$, and $\nu_i := \#\{d_j | d_j = i\}$. By Theorem 6.1, we have $\sum \mu_i \leq \alpha + A, \sum \nu_i \leq \alpha + B$ for some small integers A, B. One can classify tuples $(\mu_1, \ldots, \mu_6; \nu_2, \ldots, \nu_6)$ (resp. $(\mu_1, \ldots, \mu_5; \nu_2, \ldots, \nu_5)$) when $\alpha = 1$ (resp. $\alpha = -1$) under given conditions. It is then possible to classify initial baskets of weighted baskets with given μ_i, ν_i's. For a given weighted basket, one can compute its plurigenera by Reid's Riemann-Roch formula. By Reid's "table method", which mainly works on Poincaré series, one can determine all possible weighted complete intersections with given formal baskets.

There might be infinitely many initial baskets with given μ_i, ν_i's since the index of each individual basket might be arbitrarily large. In the Fano case with $\alpha = -1$, one can use the property $(-K_X \cdot c_2) \geq 0$ to obtain the maximal index of basket. In the case of general type, one can use $K_X^3 > 0$ to exclude most of the baskets with large indices. The details can be found in J.-J. Chen, J. A. Chen, and M. Chen [ibid.].

To summarize, the following statement holds:

Theorem 6.2. *(see J.-J. Chen, J. A. Chen, and M. Chen [ibid., Part II]) The lists of threefold weighted complete intersections in Fletcher Iano-Fletcher [2000, pp. 15.1, 15.4, 16.6, 16.7, 18.16] are complete.*

For higher dimensional weighted complete intersections, one may refer to the interesting paper of Brown and Kasprzyk [2016].

6.2 On quasi-polarized threefolds.

One can also apply the technique of baskets and singular Riemann-Roch to study some quasi-polarized threefolds (X, L). For example, C. Jiang [2016] proved the following interesting results:

- Let X be a minimal 3-fold with $K_X \equiv 0$ and L a nef and big Weil divisor. Then $|mL|$ and $|K_X + mL|$ give birational maps for all $m \geq 17$.

- Let X be a minimal Gorenstein 3-fold with $K_X \equiv 0$ and L a nef and big Weil divisor. Then $|K_X + mL|$ gives a birational map for all $m \geq 5$.

7 A brief review to explicit birational geometry of higher dimensional varieties

In dimension 4 or higher, it seems very difficult or hopeless to have desired description of terminal singularities. Therefore it is hard to study the singular Riemann-Roch formula on minimal $\mathbb{Q}$-factorial terminal n-folds. In other words, many techniques described above for 3-folds do not work in higher dimensions.

Even though little is known for dimension 4 or higher. There are some interesting results that we would like to recall here. Interested readers may find possible paths to move on.

7.1 Projective varieties with very large canonical volumes.

Apart from considering the number r_n, it is also interesting to consider another optimal constant r_n^+ so that, for all nonsingular projective n-folds X of general type with $p_g(X) > 0$, $\varphi_{m,X}$ is birational for all $m \geq r_n^+$. By definition $r_n^+ \leq r_n$ for any $n > 0$. One has $r_1^+ = r_1 = 3$ and $r_2^+ = r_2 = 5$ according to Bombieri. We start with the review of Bombieri's result:

Theorem 7.1. *(see Bombieri [1973]) Let S be a minimal surface of general type. Then*

(1) when $p_g(S) \geq 4$, $r_s(S) \leq r_1^+ = 3$;

(2) when $K_S^2 \geq 3$, $r_s(S) \leq r_1 = 3$.

The 3-dimensional analogy was realized by the second author and Todorov respectively:

Theorem 7.2. *Let X be a minimal projective 3-fold of general type. Then*

(1) when $p_g(X) \geq 4$, $r_s(X) \leq r_2^+ = 5$ (see M. Chen [2003, Theorem 1.2 (2)]);

(2) when $K_X^3 > 12^3$, $r_s(X) \leq r_2 = 5$ (see G. T. Todorov [2007] and M. Chen [2012]).

Recently the second author and Jiang proved the following analogy in dimensions 4 and 5:

Theorem 7.3. *(M. Chen and Z. Jiang [2017a, Theorem 1.4]) There exists a constant $K(4) > 0$ such that for any minimal 4-fold X with $K_X^3 > K(4)$, $r_s(X) \leq r_3$.*

Theorem 7.4. *(M. Chen and Z. Jiang [2017a, Theorem 1.5]) There exist two constants $L(4) > 0$ and $L(5) > 0$. For any minimal n-fold X of general type with $p_g(X) \geq L(n)$ ($n = 4, 5$), $r_s(X) \leq r_{n-1}^+$.*

It is natural to ask whether such analogy holds in any dimension.

7.2 Pluricanonical maps on varieties of higher Albanese dimensions.

We recall the following interesting theorem on varieties of maximal Albanese dimensions:

Theorem 7.5. *(see J. A. Chen and Hacon [2002] and Z. Jiang and Lahoz [2013]) Let X be a smooth projective variety of maximal Albanese dimension and of general type. Then $\varphi_{m,X}$ is birational for $m \geq 3$.*

The result in the above theorem is clearly optimal. In fact there is the following generalization:

Theorem 7.6. *(see Z. Jiang and Sun [2015]) Let X be a smooth projective variety of general type and of Albanese fiber dimension one. Then $\varphi_{m,X}$ is a birational map for $m \geq 4$.*

Question 7.7. Under what condition $\varphi_{3,X}$ is birational for varieties of Albanese fiber dimension one?

Can one find optimal statement for varieties of Albanese fiber dimensions 2 and 3?

7.3 Geography of varieties of general type.

It is interesting to know how those birational invariants, such as the canonical volume, p_g, $\chi(\mathcal{O})$ and so on, reflect geometric properties of a given variety. Inequalities among birational invariants play important roles in the classification theory and many other geometrical problems.

In high dimensions, the famous inequality of Yau [1977, 1978] discloses the optimal relations between c_1 and c_2 on canonically polarized varieties, namely:

$$\frac{2(n+1)}{n}|c_1^{n-2} \cdot c_2| - |c_1^n| \geq 0$$

holds for any n-dimensional canonically polarized variety. Miyaoka [1987] proved that $3c_2 - c_1^2$ is pseudo-effective for canonically quasi-polarized (i.e. K being nef and big) varieties. One notices that Greb, Kebekus, Peternell, and Taji [2015] had an interesting generalization of the inequality of Yau and Miyaoka.

For Gorenstein minimal 3-folds of general type, the following inequality was proved by the authors J. A. Chen and M. Chen [2015b] (see M. Chen [2004b] and Catanese, M. Chen, and Zhang [2006] for historical context):

$$K^3 \geq \frac{4}{3}p_g - \frac{10}{3},$$

which is sharp thanks to examples of Kobayashi [1992]. Applying a key lemma in J. A. Chen and M. Chen [2015b], Hu [2013] proved the following optimal inequality:

$$K^3 \geq \frac{4}{3}\chi(\omega) - 2.$$

In a recent work of J. A. Chen and Lai [2017], two series of examples with small "slope" in arbitrary higher dimensions were constructed.

References

X. Benveniste (1986). "Sur les applications pluricanoniques des variétés de type très général en dimension 3". *Amer. J. Math.* 108.2, pp. 433–449. MR: 833363 (cit. on p. 675).

Caucher Birkar (Mar. 2016). "Anti-pluricanonical systems on Fano varieties". arXiv: 1603.05765 (cit. on p. 672).

Caucher Birkar, Paolo Cascini, Christopher D. Hacon, and James McKernan (2010). "Existence of minimal models for varieties of log general type". *J. Amer. Math. Soc.* 23.2, pp. 405–468. MR: 2601039.

Caucher Birkar and De-Qi Zhang (2016). "Effectivity of Iitaka fibrations and pluricanonical systems of polarized pairs". *Publ. Math. Inst. Hautes Études Sci.* 123, pp. 283–331. MR: 3502099 (cit. on p. 672).

E. Bombieri (1973). "Canonical models of surfaces of general type". *Inst. Hautes Études Sci. Publ. Math.* 42, pp. 171–219. MR: 0318163 (cit. on pp. 675, 683).

Gavin Brown and Alexander Kasprzyk (2016). "Four-dimensional projective orbifold hypersurfaces". *Exp. Math.* 25.2, pp. 176–193. MR: 3463567 (cit. on p. 682).

Fabrizio Catanese, Meng Chen, and De-Qi Zhang (2006). "The Noether inequality for smooth minimal 3-folds". *Math. Res. Lett.* 13.4, pp. 653–666. MR: 2250498 (cit. on p. 685).

Hsin-Ku Chen (Oct. 2017). "Threefolds of Kodaira dimension one". arXiv: 1710.11224 (cit. on p. 681).

Jheng-Jie Chen, Jungkai A. Chen, and Meng Chen (2011). "On quasismooth weighted complete intersections". *J. Algebraic Geom.* 20.2, pp. 239–262. MR: 2762991 (cit. on p. 682).

Jungkai A. Chen and Meng Chen (2008a). "An optimal boundedness on weak $\mathbb{Q}$-Fano 3-folds". *Adv. Math.* 219.6, pp. 2086–2104. MR: 2456276 (cit. on p. 680).

– (2008b). "The canonical volume of 3-folds of general type with $\chi \leq 0$". *J. Lond. Math. Soc. (2)* 78.3, pp. 693–706. MR: 2456899 (cit. on p. 679).

– (2010a). "Explicit birational geometry of 3-folds of general type, II". *J. Differential Geom.* 86.2, pp. 237–271. MR: 2772551 (cit. on pp. 676, 677).

– (2010b). "Explicit birational geometry of threefolds of general type, I". *Ann. Sci. Éc. Norm. Supér. (4)* 43.3, pp. 365–394. MR: 2667020 (cit. on pp. 674, 676, 677).

– (2015a). "Explicit birational geometry of 3-folds and 4-folds of general type, III". *Compos. Math.* 151.6, pp. 1041–1082. MR: 3357178 (cit. on pp. 676–678).

– (2015b). "The Noether inequality for Gorenstein minimal 3-folds". *Comm. Anal. Geom.* 23.1, pp. 1–9. MR: 3291363 (cit. on p. 685).

Jungkai A. Chen, Meng Chen, and De-Qi Zhang (2007). "The 5-canonical system on 3-folds of general type". *J. Reine Angew. Math.* 603, pp. 165–181. MR: 2312557 (cit. on pp. 675, 679).

Jungkai A. Chen and Christopher D. Hacon (2002). "Linear series of irregular varieties". In: *Algebraic geometry in East Asia (Kyoto, 2001)*. World Sci. Publ., River Edge, NJ, pp. 143–153. MR: 2030451 (cit. on pp. 678, 684).

– (2007). "Pluricanonical systems on irregular 3-folds of general type". *Math. Z.* 255.2, pp. 343–355. MR: 2262735 (cit. on p. 678).

Jungkai Alfred Chen and Ching-Jui Lai (Mar. 2017). "Varieties of general type with small volumes". arXiv: 1703.03527 (cit. on p. 685).

Jungkai Chen, Meng Chen, and Zhi Jiang (2013). "On 6-canonical map of irregular threefolds of general type". *Math. Res. Lett.* 20.1, pp. 19–25. MR: 3126718.

M. Chen (2004a). "On the $\mathbb{Q}$-divisor method and its application". *J. Pure Appl. Algebra* 191.1-2, pp. 143–156. MR: 2048311 (cit. on pp. 676, 677).

Meng Chen (1998). "On pluricanonical maps for threefolds of general type". *J. Math. Soc. Japan* 50.3, pp. 615–621. MR: 1626342 (cit. on p. 675).

– (2003). "Canonical stability of 3-folds of general type with $p_g \geq 3$". *Internat. J. Math.* 14.5, pp. 515–528. MR: 1993794 (cit. on pp. 676–678, 683).

– (2004b). "Minimal threefolds of small slope and the Noether inequality for canonically polarized threefolds". *Math. Res. Lett.* 11.5-6, pp. 833–852. MR: 2106244 (cit. on p. 685).

– (2007). "A sharp lower bound for the canonical volume of 3-folds of general type". *Math. Ann.* 337.4, pp. 887–908. MR: 2285742 (cit. on pp. 676–678).

– (2011). "On anti-pluricanonical systems of $\mathbb{Q}$-Fano 3-folds". *Sci. China Math.* 54.8, pp. 1547–1560. MR: 2824958 (cit. on p. 680).

– (2012). "On an efficient induction step with $\mathrm{Nklt}(X, D)$—notes to Todorov". *Comm. Anal. Geom.* 20.4, pp. 765–779. MR: 2981839 (cit. on p. 683).

– (Apr. 2016). "On minimal 3-folds of general type with maximal pluricanonical section index". To appear in *Asian Journal of Mathematics*. arXiv: 1604.04828 (cit. on p. 678).

Meng Chen and Chen Jiang (2016). "On the anti-canonical geometry of $\mathbb{Q}$-Fano threefolds". *J. Differential Geom.* 104.1, pp. 59–109. MR: 3544286 (cit. on p. 680).

– (Aug. 2017). "On the anti-canonical geometry of weak $\mathbb{Q}$-Fano threefolds, II". arXiv: 1708.04954 (cit. on p. 680).

Meng Chen and Zhi Jiang (2017a). "A reduction of canonical stability index of 4 and 5 dimensional projective varieties with large volume". *Ann. Inst. Fourier (Grenoble)* 67 (5), pp. 2043–2082. arXiv: 1504.03939 (cit. on pp. 683, 684).

– (2017b). "A reduction of canonical stability index of 4 and 5 dimensional projective varieties with large volume". *Ann. Inst. Fourier (Grenoble)* 67.5, pp. 2043–2082. MR: 3732684.

Meng Chen and De-Qi Zhang (2008). "Characterization of the 4-canonical birationality of algebraic threefolds". *Math. Z.* 258.3, pp. 565–585. MR: 2369045 (cit. on p. 677).

Meng Chen and Kang Zuo (2008). "Complex projective 3-fold with non-negative canonical Euler-Poincaré characteristic". *Comm. Anal. Geom.* 16.1, pp. 159–182. MR: 2411471 (cit. on p. 677).

Gabriele Di Cerbo (2014). "Uniform bounds for the Iitaka fibration". *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 13.4, pp. 1133–1143. MR: 3362122 (cit. on p. 672).

Osamu Fujino and Shigefumi Mori (2000). "A canonical bundle formula". *J. Differential Geom.* 56.1, pp. 167–188. MR: 1863025 (cit. on p. 672).

Daniel Greb, Stefan Kebekus, Thomas Peternell, and Behrouz Taji (Nov. 2015). "The Miyaoka-Yau inequality and uniformisation of canonical models". arXiv: 1511.08822 (cit. on p. 684).

Christopher D. Hacon and James McKernan (2006). "Boundedness of pluricanonical maps of varieties of general type". *Invent. Math.* 166.1, pp. 1–25. MR: 2242631 (cit. on p. 672).

– (2010). "Boundedness of pluricanonical maps of varieties of general type". In: *Proceedings of the International Congress of Mathematicians, Hyderabad 2010*. Vol. 2. New Delhi: Hindustan Book Agency, pp. 427–449.

Yong Hu (Sept. 2013). "Inequality of Noether Type for Gorenstein Minimal 3-folds of General Type". arXiv: 1309.4618 (cit. on p. 685).

A. R. Iano-Fletcher (2000). "Working with weighted complete intersections". In: *Explicit birational geometry of 3-folds*. Vol. 281. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 101–173. MR: 1798982 (cit. on pp. 678, 680–682).

Chen Jiang (2016). "On birational geometry of minimal threefolds with numerically trivial canonical divisors". *Math. Ann.* 365.1-2, pp. 49–76. MR: 3498903 (cit. on p. 682).

Xiaodong Jiang (2013). "On the pluricanonical maps of varieties of intermediate Kodaira dimension". *Math. Ann.* 356.3, pp. 979–1004. MR: 3063904 (cit. on p. 672).

Zhi Jiang and Sofia Lahoz Martíand Tirabassi (2013). "On the Iitaka fibration of varieties of maximal Albanese dimension". *Int. Math. Res. Not. IMRN* 13, pp. 2984–3005. MR: 3072998 (cit. on p. 684).

Zhi Jiang and Hao Sun (2015). "Cohomological support loci of varieties of Albanese fiber dimension one". *Trans. Amer. Math. Soc.* 367.1, pp. 103–119. MR: 3271255 (cit. on p. 684).

Yujiro Kawamata (1986). "On the plurigenera of minimal algebraic 3-folds with $K \equiv 0$". *Math. Ann.* 275.4, pp. 539–546. MR: 859328 (cit. on p. 681).

– (1992). "Boundedness of **Q**-Fano threefolds". In: *Proceedings of the International Conference on Algebra, Part 3 (Novosibirsk, 1989)*. Vol. 131. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 439–445. MR: 1175897 (cit. on p. 672).

– (1996). "Divisorial contractions to 3-dimensional terminal quotient singularities". In: *Higher-dimensional complex varieties (Trento, 1994)*. de Gruyter, Berlin, pp. 241–246. MR: 1463182 (cit. on p. 675).

Yujiro Kawamata, Katsumi Matsuda, and Kenji Matsuki (1987). "Introduction to the minimal model problem". In: *Algebraic geometry, Sendai, 1985*. Vol. 10. Adv. Stud. Pure Math. North-Holland, Amsterdam, pp. 283–360. MR: 946243.

Masanori Kobayashi (1992). "On Noether's inequality for threefolds". *J. Math. Soc. Japan* 44.1, pp. 145–156. MR: 1139663 (cit. on p. 685).

János Kollár (1986). "Higher direct images of dualizing sheaves. I". *Ann. of Math. (2)* 123.1, pp. 11–42. MR: 825838 (cit. on p. 676).

János Kollár, Yoichi Miyaoka, Shigefumi Mori, and Hiromichi Takagi (2000). "Boundedness of canonical **Q**-Fano 3-folds". *Proc. Japan Acad. Ser. A Math. Sci.* 76.5, pp. 73–77. MR: 1771144 (cit. on p. 672).

János Kollár and Shigefumi Mori (1998). *Birational geometry of algebraic varieties*. Vol. 134. Cambridge Tracts in Mathematics. With the collaboration of C. H. Clemens and A. Corti, Translated from the 1998 Japanese original. Cambridge University Press, Cambridge, pp. viii+254. MR: 1658959.

Seunghun Lee (2000). "Quartic-canonical systems on canonical threefolds of index 1". *Comm. Algebra* 28.12. Special issue in honor of Robin Hartshorne, pp. 5517–5530. MR: 1808586 (cit. on p. 675).

Kenji Matsuki (1986). "On pluricanonical maps for 3-folds of general type". *J. Math. Soc. Japan* 38.2, pp. 339–359. MR: 833207 (cit. on p. 675).

Yoichi Miyaoka (1987). "The Chern classes and Kodaira dimension of a minimal variety". In: *Algebraic geometry, Sendai, 1985*. Vol. 10. Adv. Stud. Pure Math. North-Holland, Amsterdam, pp. 449–476. MR: 946247 (cit. on pp. 675, 684).

Shigefumi Mori (1988). "Flip theorem and the existence of minimal models for 3-folds". *J. Amer. Math. Soc.* 1.1, pp. 117–253. MR: 924704 (cit. on p. 673).

David R. Morrison (1986). "A remark on: "On the plurigenera of minimal algebraic 3-folds with $K \equiv 0$" [Math. Ann. **275** (1986), no. 4, 539–546; MR0859328 (88c:14049)] by Y. Kawamata". *Math. Ann.* 275.4, pp. 547–553. MR: 859329 (cit. on p. 681).

Keiji Oguiso (1993). "A remark on the global indices of **Q**-Calabi-Yau 3-folds". *Math. Proc. Cambridge Philos. Soc.* 114.3, pp. 427–429. MR: 1235989 (cit. on p. 681).

Gianluca Pacienza (2009). "On the uniformity of the Iitaka fibration". *Math. Res. Lett.* 16.4, pp. 663–681. MR: 2525032 (cit. on p. 672).

Miles Reid (1983). "Minimal models of canonical 3-folds". In: *Algebraic varieties and analytic varieties (Tokyo, 1981)*. Vol. 1. Adv. Stud. Pure Math. North-Holland, Amsterdam, pp. 131–180. MR: 715649 (cit. on p. 675).

– (1987). "Young person's guide to canonical singularities". In: *Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985)*. Vol. 46. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 345–414. MR: 927963 (cit. on p. 673).

Adam Ringler (Aug. 2007). "On a conjecture of Hacon and McKernan in dimension three". arXiv: 0708.3662 (cit. on p. 681).

Shigeharu Takayama (2006). "Pluricanonical systems on algebraic varieties of general type". *Invent. Math.* 165.3, pp. 551–587. MR: 2242627 (cit. on p. 672).

Gueorgui Tomov Todorov (2007). "Pluricanonical maps for threefolds of general type". *Ann. Inst. Fourier (Grenoble)* 57.4, pp. 1315–1330. MR: 2339333 (cit. on p. 683).

Gueorgui Todorov and Chenyang Xu (2009). "Effectiveness of the log Iitaka fibration for 3-folds and 4-folds". *Algebra Number Theory* 3.6, pp. 697–710. MR: 2579391 (cit. on p. 672).

Hajime Tsuji (2006). "Pluricanonical systems of projective varieties of general type. I". *Osaka J. Math.* 43.4, pp. 967–995. MR: 2303558 (cit. on p. 672).

Eckart Viehweg and De-Qi Zhang (2009). "Effective Iitaka fibrations". *J. Algebraic Geom.* 18.4, pp. 711–730. MR: 2524596 (cit. on p. 672).

P. M. H. Wilson (1980). "The pluricanonical map on varieties of general type". *Bull. London Math. Soc.* 12.2, pp. 103–107. MR: 571729 (cit. on p. 675).

Shing Tung Yau (1977). "Calabi's conjecture and some new results in algebraic geometry". *Proc. Nat. Acad. Sci. U.S.A.* 74.5, pp. 1798–1799. MR: 0451180 (cit. on p. 684).

– (1978). "On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I". *Comm. Pure Appl. Math.* 31.3, pp. 339–411. MR: 480350 (cit. on p. 684).

Received 2017-12-10.

JUNGKAI A. CHEN (陳榮凱)

National Center for Theoretical Sciences
Taipei, 106, Taiwan

and

Department of Mathematics
National Taiwan University
Taipei, 106, Taiwan
jkchen@ntu.edu.tw

Meng Chen (陈猛)
School of Mathematical Sciences & Shanghai Centre for Mathematical Sciences
Fudan University
Shanghai 200433, China
mchen@fudan.edu.cn

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (689–716)

MIRROR SYMMETRY AND CLUSTER ALGEBRAS

PAUL HACKING AND SEAN KEEL

Abstract

We explain our proof, joint with Mark Gross and Maxim Kontsevich, of conjectures of Fomin–Zelevinsky and Fock–Goncharov on canonical bases of cluster algebras. We interpret a cluster algebra as the ring of global functions on a non-compact Calabi–Yau variety obtained from a toric variety by a blow up construction. We describe a canonical basis of a cluster algebra determined by tropical counts of holomorphic discs on the mirror variety, using the algebraic approach to the Strominger–Yau–Zaslow conjecture due to Gross and Siebert.

1 Introduction

We say a complex variety U is *log Calabi–Yau* if it admits a smooth projective compactification X with normal crossing boundary[1] D such that $K_X + D = 0$, that is, there is a nowhere zero holomorphic top form Ω on U with simple poles along D. The mirror symmetry phenomenon for compact Calabi–Yau manifolds extends to the case of log Calabi–Yau varieties, see Auroux [2009] and Section 4. We say U has *maximal boundary* if D has a 0-stratum (a point cut out by $n = \dim_{\mathbb{C}} X$ branches of D) and *positive* if D is the support of an ample divisor[2] (so in particular U is affine) . The *tropical set* $U^{\mathrm{trop}}(\mathbb{R})$ of U is the cone over the dual complex of D; we write $U^{\mathrm{trop}}(\mathbb{Z})$ for its integral points.

Conjecture 1-1. *Mirror symmetry defines an involution on the set of positive log Calabi–Yau varieties with maximal boundary. For a mirror pair U and V, there is a basis ϑ_q, $q \in U^{\mathrm{trop}}(\mathbb{Z})$ of $H^0(V, \mathcal{O}_V)$ parametrized by the integral points of the tropical set of U, which is canonically determined up to multiplication by scalars $\lambda_q \in \mathbb{C}^\times$, $q \in U^{\mathrm{trop}}(\mathbb{Z})$.*

This paper is based on joint work with Mark Gross and Maxim Kontsevich. The algebraic approach to the Strominger–Yau–Zaslow conjecture is due to Gross and Siebert. The authors were partially supported by NSF grants DMS-1601065 (P.H.) and DMS-1561632 (S.K.).
MSC2010: primary 14J33; secondary 13F60.

[1] More generally, (X, D) has $\mathbb{Q}$-factorial divisorial log terminal singularities (Kollár and Mori [1998], Definition 2.37).

[2] More generally, D is the support of a big and nef divisor.

For example, if $U \simeq (\mathbb{C}^\times)^n$ is an algebraic torus, then the mirror V is the dual algebraic torus, and the canonical basis is given by the characters of V (up to scalars), which may be characterized as the units of $H^0(V, \mathcal{O}_V)$. The set of characters of V corresponds under the duality to the set of 1-parameter subgroups of U, which is identified with $U^{\text{trop}}(\mathbb{Z})$. The heuristic justification for Conjecture 1-1 coming from mirror symmetry is explained in Section 4.

Cluster algebras were introduced by Fomin and Zelevinsky as a tool to understand the constructions of canonical bases in representation theory by Lusztig Fomin and Zelevinsky [2002]. In Section 2 we review a description of cluster varieties in terms of toric and birational geometry Gross, Hacking, and Keel [2015a]. Roughly speaking, a cluster variety is a log Calabi–Yau variety U which carries a non-degenerate holomorphic 2-form and is obtained from a toric variety $\bar{X}$ by blowing up codimension 2 centers in the toric boundary and removing its strict transform. The existence of the 2-form greatly constrains the possible centers and accounts for the combinatorial description of cluster varieties. The mutations of cluster theory are given by elementary transformations of $\mathbb{P}^1$-bundles linking different toric models.

For a cluster variety U, Fock and Goncharov defined a dual cluster variety V by an explicit combinatorial recipe, and stated the analogue of Conjecture 1-1 in this setting Fock and Goncharov [2006]. In Section 5 we use an algebraic version of the Strominger–Yau–Zaslow mirror construction Strominger, Yau, and Zaslow [1996] to explain that if U is positive then V should be its mirror. (If U is not positive, then we expect that the mirror of U is an open analytic subset of V and the Fock–Goncharov conjecture is false, cf. Gross, Hacking, and Keel [2015a].) Under a hypothesis on U related to positivity, our construction proves Conjecture 1-1 in this case. In particular, the hypothesis is satisfied in the case of the mirror of the base affine space G/N for $G = \mathrm{SL}_m$ studied by Fomin and Zelevinsky, so we obtain canonical bases of representations of G by the Borel–Weil–Bott theorem.

2 Log Calabi–Yau varieties

Definition 2-1. A *log Calabi–Yau pair* (X, D) is a smooth complex projective variety X together with a reduced normal crossing divisor $D \subset X$ such that $K_X + D = 0$. Thus there is a nowhere zero holomorphic top form Ω on $U = X \setminus D$ (a *holomorphic volume form*) such that Ω has a simple pole along each component of D, uniquely determined up to multiplication by a nonzero scalar.

We say a variety U is *log Calabi–Yau* if there exists a log Calabi–Yau pair (X, D) such that $U = X \setminus D$.

Remark 2-2. Note that if U is a smooth variety and (X, D) is any normal crossing compactification of U, the subspace $H^0(\Omega^p_X(\log D)) \subset H^0(\Omega^p_U)$ for each $p \geq 0$ is independent of (X, D) Deligne [1971]. In particular, if U is a log Calabi–Yau variety then there is a holomorphic volume form Ω on U such that Ω has at worst a simple pole along each boundary divisor of any normal crossing compactification (X, D), uniquely determined up to a scalar.

Definition 2-3. We say a log Calabi–Yau pair (X, D) has *maximal boundary* if the boundary D has a 0-stratum, that is, a point $p \in D \subset Y$ cut out by $n = \dim_{\mathbb{C}} X$ analytic branches of the divisor D, so that we have a local analytic isomorphism

$$(p \in D \subset X) \simeq (0 \in (z_1 \cdots z_n = 0) \subset \mathbb{C}^n).$$

We say a log Calabi–Yau variety U has maximal boundary if some (equivalently, any de Fernex, Kollár, and Xu [2012], Proposition 11) log Calabi–Yau compactification (X, D) of U has maximal boundary.

Definition 2-4. We say a log Calabi–Yau variety U is *positive* if there exists a log Calabi–Yau compactification $(X, D = \sum D_i)$ and positive integers a_i such that $A = \sum a_i D_i$ is ample. In particular, $U = X \setminus D$ is affine.

Example 2-5. The algebraic torus $(\mathbb{C}^\times)^n$ is a log Calabi-Yau variety, with holomorphic volume form $\Omega = \frac{dz_1}{z_1} \wedge \cdots \wedge \frac{dz_n}{z_n}$. Any toric compactification (X, D) satisfies $K_X + D = 0$.

Example 2-6 (Non-toric blow up)**.** Let (X, D) be a log Calabi–Yau pair and $Z \subset X$ a smooth subvariety of codimension 2 which is contained in a unique component of D and meets the other components transversely. Let $\pi : \tilde{X} \to X$ be the blow up of Z and $\tilde{D} \subset \tilde{X}$ the strict transform of D. Then the pair $(\tilde{X}, \tilde{D})$ is log Calabi–Yau.

Definition 2-7. A *toric model* of a log Calabi–Yau variety U is a log Calabi–Yau compactification (X, D) of U together with a birational morphism $f : (X, D) \to (\bar{X}, \bar{D})$ such that $(\bar{X}, \bar{D})$ is a toric variety together with its toric boundary and f is a composition of non-toric blow ups as in Example 2-6.

Remark 2-8. In the description of a log Calabi–Yau variety U in terms of a toric model $(X, D) \to (\bar{X}, \bar{D})$, one can replace the projective toric variety $\bar{X}$ with the toric open subset $\bar{X}' \subset \bar{X}$ given by the union of the big torus $T \subset \bar{X}$ and the open T-orbit in each boundary divisor containing the center of one of the blow ups. Thus the fan Σ' of $\bar{X}'$ is the subset of the fan of $\bar{X}$ consisting of $\{0\}$ and the rays corresponding to these boundary divisors. Cf. Gross, Hacking, and Keel [2015a], §3.2.

Example 2-9. Let $X = \mathbb{P}^2$ and $D = Q + L$ the union of a smooth conic Q and a line L meeting transversely. We describe a toric model of the log Calabi–Yau surface $U = X \setminus D$. First, choose a point $p \in Q \cap L$ and blow up at p. Second, blow up at the intersection point of the exceptional divisor and the strict transform of Q. Let $\tilde{X} \to X$ be the composition of the two blow ups and $\tilde{D} = \pi^{-1}D$. The strict transform in $\tilde{X}$ of the tangent line to Q at p is a (-1)-curve E meeting $\tilde{D}$ transversely at a point of the exceptional divisor of the second blow up. Contracting E yields a toric pair $(\bar{X}, \bar{D})$ with $\bar{X} \simeq \mathbb{F}_2 := \mathbb{P}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-2))$.

Proposition 2-10. *(Gross, Hacking, and Keel [2015b], Proposition 1.3) Let U be a log Calabi-Yau surface with maximal boundary. Then U has a toric model.*

The proof is an exercise in the birational geometry of surfaces.

Example 2-11. We describe an example of a log Calabi–Yau 3-fold with maximal boundary which is irrational. In particular, it does not have a toric model.

Smooth quartic 3-folds are irrational Iskovskih and Manin [1971]. Let $X \subset \mathbb{P}^4$ be a smooth quartic 3-fold with hyperplane section

$$D = (X_1^4 + X_2^4 + X_3^4 + X_1X_2X_3X_4 = 0) \subset \mathbb{P}^3.$$

The surface D has a unique singular point $p = (0 : 0 : 0 : 1)$. The minimal resolution of D is obtained by blowing up p and has exceptional locus a triangle of (-3)-curves (in particular, $p \in D$ is a *cusp singularity*).

We describe a sequence of blow ups $\pi : \tilde{X} \to X$ such that the inverse image of D is a normal crossing divisor. First blow up the point p. The inverse image D^1 of D consists of two components, the exceptional divisor $E \simeq \mathbb{P}^2$ and the strict transform D' of D (which is its minimal resolution). The intersection $E \cap D'$ is the exceptional locus of $D' \to D$, a triangle of smooth rational curves. Blow up the nodes of the triangle. For each of the exceptional divisors $E_i \simeq \mathbb{P}^2$, the strict transforms of E and D meet E_i in a common line $l_i \subset E_i$. Finally blow up each line l_i to obtain $\tilde{X}$.

Define the divisor $\tilde{D} \subset \tilde{Y}$ to be the union of the strict transforms of D, E, and the exceptional divisors over the lines l_i. Then $K_{\tilde{X}} + \tilde{D} = 0$. The variety $\tilde{U} = \tilde{X} \setminus \tilde{D}$ is an irrational log Calabi–Yau 3-fold with maximal boundary.

Remark 2-12. On the other hand, if (X, D) is a log Calabi-Yau pair with maximal boundary then X is rationally connected Kollár and Xu [2016], (18).

3 Cluster varieties

3.1 Birational geometric description of cluster varieties.

Definition 3-1. We say a log Calabi–Yau variety U is a *cluster variety* if

1. There is a non-degenerate holomorphic 2-form σ on U such that for some (equivalently, any Deligne [1971]) normal crossing compactification (X, D) we have $\sigma \in H^0(\Omega^2_X(\log D))$.

2. U has a toric model.

Remark 3-2. It is customary in the theory of cluster algebras to allow the 2-form to be degenerate. However, the non-degenerate case is the essential one (cf. Section 5.4).

Example 3-3. Every log Calabi–Yau surface with maximal boundary is a cluster variety by Proposition 2-10.

Suppose U is a cluster variety with 2-form σ and toric model $f : (X, D) \to (\bar{X}, \bar{D})$. Then $\sigma = f^*\bar{\sigma}$ for some $\bar{\sigma} \in H^0(\Omega^2_{\bar{X}}(\log \bar{D}))$ by Hartogs' theorem. The sheaf $\Omega_{\bar{X}}(\log D)$ is freely generated by $\frac{dz_1}{z_1}, \ldots, \frac{dz_n}{z_n}$, where $z_1, \ldots, z_n$ is a basis of characters for the algebraic torus $T = \bar{X} \setminus D \simeq (\mathbb{C}^\times)^n_{z_1,\ldots,z_n}$. See Fulton [1993], Proposition, p. 87. Thus $\bar{\sigma} = \frac{1}{2}\sum a_{ij}\frac{dz_i}{z_i} \wedge \frac{dz_j}{z_j}$ for some non-degenerate skew matrix (a_{ij}). The following lemma is left as an exercise.

Lemma 3-4. *Let (X, D) be a normal crossing pair, $Z \subset X$ a smooth codimension 2 subvariety contained in a unique component F of D and meeting the remaining components transversely, $\pi : \tilde{X} \to X$ the blow up with center Z, and $\tilde{D}$ the strict transform of D. Let $\sigma \in H^0(\Omega^2_X(\log D))$ be a log 2-form on X. Let $D_F = (D - F)|_F$ and let $\mathrm{Res}_F : \Omega^2_X(\log D) \to \Omega_F(\log D_F)$ be the Poincaré residue map. Then σ lifts to a log 2-form on $(\tilde{X}, \tilde{D})$ if and only if $(\mathrm{Res}_F\, \sigma)|_Z = 0$.*

Now let $Z \subset F \subset \bar{D}$ be the center of one of the blow ups for the toric model f. We may choose coordinates on T so that $F \setminus \bar{D}_F = (z_1 = 0) \subset \mathbb{A}^1_{z_1} \times (\mathbb{C}^\times)^{n-1}_{z_2,\ldots,z_n}$, then $\mathrm{Res}_F(\bar{\sigma}) = \sum_{j>1} a_{1j}\frac{dz_j}{z_j}$. Using the lemma, we deduce that

1. $\mathrm{Res}_F(\bar{\sigma})$ is proportional to an integral log 1-form, that is
$$\mathrm{Res}_F(\bar{\sigma}) = \nu \cdot \sum_{j>1} b_j \frac{dz_j}{z_j}$$
for some $\nu \in \mathbb{C}^\times$ and pairwise coprime $b_j \in \mathbb{Z}$. Equivalently, writing $\chi : T \to \mathbb{C}^\times$ for the character $\prod z_j^{b_j}$, $\mathrm{Res}_F(\bar{\sigma}) = \nu \frac{d\chi}{\chi}$.

2. $Z = F \cap \overline{(\chi = \lambda)}$ for some $\lambda \in \mathbb{C}^\times$.

Remark 3-5. Note that, after a change of coordinates, we may assume $\chi = z_2$. Thus, if f is a single blow up, then $U = X \setminus D$ decomposes as a product $U' \times (\mathbb{C}^\times)^{n-2}_{z_3,\ldots,z_n}$. In general U does not globally decompose as a product.

Conversely, any sequence of non-toric blow ups of $(\bar{X}, \bar{D}, \bar{\sigma})$ with the above properties yields a cluster variety.

3.2 Atlas of tori and elementary transformations.

The usual description of a cluster variety U is as follows: The variety U is the union of a countable collection of open subsets T_α (indexed by *seeds* α) which are copies of a fixed algebraic torus $T \simeq (\mathbb{C}^\times)^n$. The glueing maps between the open subsets are compositions of *mutations*, given (for some choice of coordinates $z_1, \ldots, z_n$ on T) by the formula

$$\mu \colon T \dashrightarrow T', \quad (z_1, z_2, \ldots, z_n) \mapsto (z_1(1 + cz_2)^{-1}, z_2, z_3, \ldots, z_n)$$

for some $c \in \mathbb{C}^\times$.

There is the following geometric interpretation. First, note that a toric model $f \colon (X, D) \to (\bar{X}, \bar{D})$ determines an open inclusion of the torus $T = \bar{X} \setminus \bar{D}$ in $U = X \setminus D$ via f^{-1}. This is the origin of the torus charts of a cluster variety: seeds correspond to toric models. Second, mutations correspond to birational transformations between toric models given by elementary transformations of $\mathbb{P}^1$-bundles. In the above notation, let $(\bar{X}, \bar{D})$ be the toric partial compactification of T given by $\mathbb{P}^1_{z_1} \times (\mathbb{C}^\times)^{n-1}_{z_2,\ldots,z_n}$. Let $Z = (z_1 = 0) \cap (1 + cz_2 = 0)$, let $\pi \colon X \to \bar{X}$ be the blow up of Z, and D the strict transform of $\bar{D}$. Write $H = (1 + cz_2 = 0) \subset \bar{X}$ and let $H' \subset X$ be its strict transform. Then H' can be blown down, yielding a morphism $\pi' \colon (X, D) \to (\bar{X}', \bar{D}')$ to a second toric pair such that $\bar{X}'$ is also isomorphic to $\mathbb{P}^1_{z_1} \times (\mathbb{C}^\times)^{n-1}_{z_2,\ldots,z_n}$ and π' is the blow up of $Z' = (z_1 = \infty) \cap (1 + cz_2 = 0)$. The birational map $(\bar{X}, \bar{D}) \dashrightarrow (\bar{X}', \bar{D}')$ is an *elementary transformation* of $\mathbb{P}^1$-bundles over $(\mathbb{C}^\times)^{n-1}$. Writing $U = X \setminus D$, $T = \bar{X} \setminus \bar{D}$, and $T' = \bar{X}' \setminus \bar{D}'$, we have $T \cup T' = U \setminus W$ where $W \simeq Z \simeq Z'$ is the intersection of the exceptional divisors of π and π'. The mutation $\mu \colon T \dashrightarrow T'$ is the restriction of the birational map $\bar{X} \dashrightarrow \bar{X}'$.

Remark 3-6. In general the union of tori in the original definition of a cluster variety is an open subset of a cluster variety in the sense of Definition 3-1 with complement of codimension at least 2 provided that the parameters $\lambda \in \mathbb{C}^\times$ are very general Gross, Hacking, and Keel [2015a], Theorem 3.9. For simplicity we will always assume that this is the case.

3.3 Combinatorial data for toric model of a cluster variety.

We can give an intrinsic description of the data for a toric model of a cluster variety as follows. (We use the notation of Fulton [1993] for toric varieties.) Let $T = \bar{X} \setminus \bar{D}$ be the big torus acting on $\bar{X}$. Let $N = H_1(T, \mathbb{Z}) = \operatorname{Hom}(\mathbb{C}^\times, T)$ be the lattice of 1-parameter subgroups of T and $M = N^* = \operatorname{Hom}(T, \mathbb{C}^\times)$ the dual lattice of characters of T. Then $T = N \otimes_{\mathbb{Z}} \mathbb{C}^\times$. We sometimes use the multiplicative notation z^m for characters.

Let U be a cluster variety and $f : (X, D) \to (\bar{X}, \bar{D})$ be a toric model for U. With notation as in Section 3.1, let $Z = F \cap \overline{(\chi = \lambda)}$ be the center of one of the blow ups. The toric boundary divisor $F \subset \bar{D}$ corresponds to a primitive vector $v \in N$ (the generator of the corresponding ray of the fan of $\bar{X}$). The character χ corresponds to a primitive element $m \in v^\perp \subset M$ (primitive because Z is assumed irreducible). The 2-form $\bar{\sigma}$ lies in $H^0(\Omega^2_{\bar{X}}(\log \bar{D})) = \wedge^2 M_{\mathbb{C}}$. The condition $\operatorname{Res}_F(\bar{\sigma}) = \nu \cdot \frac{d\chi}{\chi}$ is equivalent to $\bar{\sigma}(v, \cdot) = \nu \cdot m$. The associated mutation is given by

$$\text{(3-7)} \qquad \mu = \mu_{(m,v)} : T \dashrightarrow T, \quad \mu^*(z^{m'}) = z^{m'}(1 + cz^m)^{-\langle m', v\rangle}$$

where $c = -1/\lambda$.

3.4 The tropicalization of a log Calabi–Yau variety.

Definition 3-8. Let U be a log Calabi–Yau variety. We define the *tropicalization* $U^{\text{trop}}(\mathbb{R})$ of U as follows. Let (X, D) be a log Calabi–Yau compactification of U. We may assume (blowing up boundary strata if necessary) that D is a simple normal crossing divisor, that is, each component D_i of D is smooth and each intersection $D_{i_1} \cap \cdots \cap D_{i_k}$ is either irreducible or empty. The *dual complex* of D is the simplicial complex with vertex set indexed by components of D, such that a set of vertices spans a simplex if and only if the intersection of the corresponding divisors is non-empty. Let $U^{\text{trop}}(\mathbb{R})$ be the cone over the dual complex of D, and $U^{\text{trop}}(\mathbb{Z})$ its integral points. One can show using Abramovich, Karu, Matsuki, and Włodarczyk [2002] that $U^{\text{trop}}(\mathbb{R})$ is independent of the choice of (X, D) up to $\mathbb{Z}$PL-homeomorphism Kontsevich and Soibelman [2006, Sec. 6.6], de Fernex, Kollár, and Xu [2012], Proposition 11.

The set $U^{\text{trop}}(\mathbb{Z})$ has the following intrinsic description: Let Ω be a holomorphic volume form on U as in Remark 2-2. Then $U^{\text{trop}}(\mathbb{Z}) \setminus \{0\}$ is identified with the set of pairs (ν, k) consisting of a divisorial valuation $\nu : \mathbb{C}(U)^\times \to \mathbb{Z}$ such that $\nu(\Omega) < 0$ and a positive integer k. Thus roughly speaking $U^{\text{trop}}(\mathbb{Z}) \setminus \{0\}$ is the set of all pairs (F, k) where F is a boundary divisor in some log Calabi-Yau compactification (X, D) of U and $k \in \mathbb{N}$.

Example 3-9. If $U = T = N \otimes \mathbb{C}^\times$ is an algebraic torus then we have an identification $U^{\text{trop}}(\mathbb{Z}) = N$: Given $0 \neq v \in N$ write $v = kv'$ where $k \in \mathbb{N}$ and $v' \in N$ is primitive. Then v' corresponds to a toric boundary divisor associated to the ray $\rho = \mathbb{R}_{\geq 0} \cdot v'$ in $N_{\mathbb{R}}$, with associated valuation $\nu : \mathbb{C}(T)^\times \to \mathbb{Z}$ determined by $\nu(z^m) = \langle m, v'\rangle$. These are the only divisors along which Ω has a pole, by Kollár and Mori [1998], Lemmas 2.29 and 2.45.

If (X, D) is a toric compactification of T then the cone over the dual complex of D is identified with the fan Σ of X in $U^{\mathrm{trop}}(\mathbb{R}) = N_{\mathbb{R}}$.

If f is a nonzero rational function on a log Calabi–Yau variety U, then we have a $\mathbb{Z}$PL map $f^{\mathrm{trop}}\colon U^{\mathrm{trop}}(\mathbb{R}) \to \mathbb{R}$ defined on primitive integral points $v = (\nu, 1)$ by $f^{\mathrm{trop}}(v) = \nu(f)$. If $f\colon U \dashrightarrow V$ is a birational map between log Calabi–Yau varieties, then there is a canonical $\mathbb{Z}$PL identification $f^{\mathrm{trop}}\colon U^{\mathrm{trop}}(\mathbb{R}) \to V^{\mathrm{trop}}(\mathbb{R})$ defined by $f^{\mathrm{trop}}(\nu) = \nu \circ f^*$.

Example 3-10. For the mutation (3-7), we have

$$\mu^{\mathrm{trop}}\colon N_{\mathbb{R}} \to N_{\mathbb{R}}, \quad \mu^{\mathrm{trop}}(w) = \begin{cases} w & \text{if } \langle m, w\rangle \geq 0 \\ w - \langle m, w\rangle v & \text{if } \langle m, w\rangle < 0. \end{cases}$$

4 Mirror symmetry

Mirror symmetry is a phenomenon arising in theoretical physics which predicts that Calabi–Yau varieties (together with a choice of Kähler form) come in mirror pairs U and V such that the symplectic geometry of U is equivalent to the complex geometry of V, and vice versa.

4.1 The Strominger–Yau–Zaslow conjecture. Recall that a submanifold L of a symplectic manifold (U, ω) is *Lagrangian* if $\dim_{\mathbb{R}} L = \frac{1}{2}\dim_{\mathbb{R}} U$ and $\omega|_L = 0$. Let U be a log Calabi–Yau manifold with holomorphic volume form Ω and Kähler form ω. We say a Lagrangian submanifold L of (U, ω) is *special Lagrangian* if $\operatorname{Im}\Omega|_L = 0$. The Strominger–Yau–Zaslow conjecture asserts that mirror Calabi-Yau varieties admit dual special Lagrangian torus fibrations Strominger, Yau, and Zaslow [1996]. More precisely, there exist continuous maps $f\colon U \to B$ and $g\colon V \to B$ with common base B and a dense open set $B^o \subset B$ such that

1. The restrictions $f^o\colon U^o \to B^o$ and $g^o\colon V^o \to B^o$ are C^∞ real n-torus fibrations such that the fibers are special Lagrangian, and
2. The associated local systems $R^1 f^o_* \mathbb{Z}$ and $R^1 g^o_* \mathbb{Z}$ on B^o are dual.

Example 4-1. Let $U = (\mathbb{C}^\times)^n_{z_1,\dots,z_n}$,

$$\Omega = \left(\frac{1}{2\pi i}\right)^n \frac{dz_1}{z_1} \wedge \cdots \wedge \frac{dz_n}{z_n} \qquad \text{and} \qquad \omega = \frac{1}{2\pi}\frac{i}{2}\sum_{j=1}^{n} \frac{dz_j}{z_j} \wedge \frac{d\bar{z}_j}{\bar{z}_j}$$

Then the map $f: U \to \mathbb{R}^n$, $f(z_1, \dots, z_n) = (\log|z_1|, \dots, \log|z_n|)$ is a special Lagrangian torus fibration. (Topologically f is the quotient by the compact torus $(S^1)^n \subset (\mathbb{C}^\times)^n$.)

Example 4-2. Let $U = T = N \otimes \mathbb{C}^\times \simeq (\mathbb{C}^\times)^n$ and let (X, D) be a smooth projective toric compactification. Let A be an ample line bundle on X. Let $K \subset T$ be the compact torus. Then, using the description of X as a GIT quotient of affine space, the Kempf–Ness theorem (Mumford, Fogarty, and Kirwan [1994], Theorem 8.3), and symplectic reduction, one can construct a K-invariant Kähler form ω on X in class $c_1(A) \in H^2(X, \mathbb{R})$. If μ is the associated moment map, then μ maps X onto the lattice polytope $P \subset M_{\mathbb{R}}$ associated to (X, A), and is topologically the quotient by K. The restriction of μ to T is a special Lagrangian torus fibration for the Kähler form $\omega|_U$ and holomorphic volume form Ω as in Example 4-1.

Construction 4-3. If $f: (U, \omega) \to B$ is a Lagrangian torus fibration, then the locus $B^o \subset B$ of smooth fibers inherits an *integral affine structure* (an atlas of charts with transition functions of the form $\mathbf{x} \mapsto A\mathbf{x} + \mathbf{b}$ for some $A \in \mathrm{GL}(n, \mathbb{Z})$ and $\mathbf{b} \in \mathbb{R}^n$). This may be constructed as follows. Fix $b_0 \in B^o$ and let $W \subset B^o$ be a small contractible neighborhood of b_0. For $\gamma \in H_1(f^{-1}(b_0), \mathbb{Z})$ define $y_\gamma: W \to \mathbb{R}$ by $y_\gamma(b) = \int_\Gamma \omega$ where $\Gamma \subset X$ is a cylinder fibering over a path from b_0 to b in W swept out by a loop in the class γ. Applying this construction to a basis of $H_1(f^{-1}(b_0), \mathbb{Z}) \simeq \mathbb{Z}^n$ gives a system of integral affine coordinates $y_1, \dots, y_n$ on $W \subset B^o$. In Examples 4-1 and 4-2 this integral affine structure is the restriction of the standard integral affine structure on $\mathbb{R}^n$.

Example 4-4. Let U, (X, D), etc. be as in Example 4-2 and consider a non-toric blow up $(\tilde{X}, \tilde{D})$ of (X, D) as in Example 2-6. Then we can modify the moment map $\mu: X \to P$ to obtain a map $\tilde{\mu}: \tilde{X} \to \tilde{P}$ such that the restriction $f: \tilde{U} \to B$ to the interior B of $\tilde{P}$ is a Lagrangian torus fibration with singular fibers Abouzaid, Auroux, and Katzarkov [2016].

Assume first that $n = \dim_{\mathbb{C}} X = 2$. Thus we have a smooth point p of D, $\pi: \tilde{X} \to X$ is the blow up of p, and $\tilde{D} \subset \tilde{X}$ is the strict transform of D. Let $S^1 \subset K$ be the stabilizer of the point $p \in X$, so that S^1 acts on $\tilde{X}$. Let $e \subset P$ be the edge of P containing $\mu(p)$, and choose integral affine coordinates y_1, y_2 on $M_{\mathbb{R}} \simeq \mathbb{R}^2$ such that $p = (0, 0)$, $e \subset (y_1 = 0)$, and $P \subset (y_1 \geq 0)$. Let $\epsilon > 0$ be sufficiently small so that the triangle T with vertices $(0, -\epsilon/2), (0, +\epsilon/2), (\epsilon, 0)$ is contained in P and its intersection with the boundary of P is contained in the interior of e. Let $\tilde{P}$ be the topological space obtained by collapsing $T \subset P$ via the map $y_1: T \to [0, \epsilon]$. Then $\tilde{P}$ is the base of an S^1-invariant map $\tilde{\mu}: (\tilde{X}, \tilde{\omega}) \to \tilde{P}$ such that the restriction to the interior B of $\tilde{P}$ is a Lagrangian torus fibration with a unique singular fiber (a pinched torus) over the image $q \in B$ of the point $(\epsilon, 0) \in P$. The fibration has monodromy around q given by the Dehn twist in the vanishing cycle (the class of the S^1-orbits). The exceptional (-1)-curve $E \subset \tilde{X}$ fibers over the interval $I \subset \tilde{P}$ given by the image of $T \subset P$. The class of the symplectic form in $H^2(\tilde{X}, \mathbb{R})$ is $[\tilde{\omega}] = \pi^*[\omega] - \epsilon c_1(E) = c_1(\pi^* A - \epsilon E)$. The symplectic form $\tilde{\omega}$ and the fibration $\tilde{\mu}$ agree with ω and μ over the complement of a tubular neighborhood of $I \subset \tilde{P}$.

A similar construction applies in dimension $n > 2$ Abouzaid, Auroux, and Katzarkov [2016], §4. (Here we work over the open set $X' \subset X$ given by the complement of the codimension two strata, cf. Remark 2-8.) Applying this construction repeatedly, we can construct a Lagrangian torus fibration on any log Calabi-Yau variety with a toric model. (Note that the fibration is *not* special Lagrangian (but cf. Abouzaid, Auroux, and Katzarkov [ibid.], Remark 4.6).)

4.2 Homological mirror symmetry. The homological mirror symmetry conjecture of Kontsevich [1995] asserts the following mathematical formulation of mirror symmetry: For mirror compact Calabi–Yau varieties U and V, the derived Fukaya category $\mathfrak{F}(U)$ of U is equivalent to the derived category of coherent sheaves $D(V)$ on V. Roughly speaking, the objects of the Fukaya category of U are Lagrangian submanifolds L together with a unitary local system, and the morphisms are given by Lagrangian Floer cohomology. See Auroux [2014] for an introduction.

If U is a log Calabi–Yau variety then, at least if U is positive (Definition 2-4), the HMS conjecture is expected to hold with the following adjustments. First, we must allow non-compact Lagrangian submanifolds with controlled behaviour at infinity. Second, the definition of the morphisms in the Fukaya category is modified at infinity using a Hamiltonian vector field associated to a function $H : U \to \mathbb{R}$ such that $H \to \infty$ sufficiently fast at infinity. The resulting category $\mathfrak{F}(U)$ is called the wrapped Fukaya category Auroux [ibid.], §4.

The HMS and SYZ conjectures are related as follows. Suppose U and V are mirror log Calabi–Yau varieties with dual Lagrangian torus fibrations $f : U \to B$ and $g : V \to B$. Let $L = f^{-1}(b)$ be a smooth fiber of f. The rank 1 unitary local systems ∇ on L are classified by their holonomy $\mathrm{hol}(\nabla) \in \mathrm{Hom}(\pi_1(L), U(1)) = L^*$ (the dual torus). It is expected that the pairs $[(L, \nabla)] \in \mathfrak{F}(U)$ correspond under the equivalence $\mathfrak{F}(U) \simeq D(V)$ to the skyscraper sheaves $\mathcal{O}_p \in D(V)$ for $p \in g^{-1}(b) \simeq L^*$. (More generally, the equivalence should be given by a real version of the relative Fourier–Mukai transform for the dual torus fibrations, cf. Kontsevich and Soibelman [2001], §9, Polishchuk [2003], §6.)

It follows that, to a first approximation, one can regard the mirror V of U as the moduli space of pairs $[(L, \nabla)]$ where L is a fiber of f and ∇ is a $U(1)$ local system. We define local holomorphic coordinates on V as follows (a complexified version of the integral affine coordinates y_γ of Construction 4-3). For $L_0 = f^{-1}(b_0)$ a smooth fiber, $\gamma \in$

$H_1(L_0, \mathbb{Z})$, and $(L = f^{-1}(b), \nabla)$ a nearby fiber together with a $U(1)$ local system, let Γ be a cylinder over a short path from b_0 to b with initial fiber Γ_{b_0} in class γ and final fiber Γ_b. We define

$$z^\gamma([(L, \nabla)]) = \exp\left(-2\pi y_\gamma(b)\right) \cdot \operatorname{hol}_\nabla(\Gamma_b) = \exp\left(-2\pi \int_\Gamma \omega\right) \cdot \operatorname{hol}_\nabla(\Gamma_b).$$

Suppose now that U is a log Calabi–Yau variety, and (X, D) is a log Calabi–Yau compactification such that ω extends to a 2-form on X. The homology groups $H_2(X, L = f^{-1}(b))$ form a local system over B^o; let $T\colon H_2(X, L_0) \to H_2(X, L)$ be the local trivialization given by parallel transport. Then, for $\beta \in H_2(X, L_0)$, we define

$$z^\beta([(L, \nabla)]) = \exp\left(-2\pi \int_{T(\beta)} \omega\right) \cdot \operatorname{hol}_\nabla(\partial T(\beta)).$$

Then $z^\beta = c z^{\partial\beta}$ where $c = \exp(-2\pi \int_\beta \omega) \in \mathbb{R}_{>0}$.

We can attempt to define global holomorphic functions ϑ_q on V for each $q = (F, k) \in U^{\mathrm{trop}}(\mathbb{Z}) \setminus \{0\}$ as follows Cho and Oh [2006], Auroux [2009]. Let (X, D) be a log Calabi–Yau compactification of U such that F is a component of the boundary D and ω extends to X. Let L be a smooth fiber of f. For $\beta \in H_2(X, L, \mathbb{Z})$, let N_β be the (virtual) count of holomorphic discs $h\colon (\mathbb{D}, \partial\mathbb{D}) \to (X, L)$ such that h meets F with contact order k and is disjoint from the remaining boundary divisors, and $h(\partial\mathbb{D})$ passes through a general point $p \in L$. We assume that N_β is well defined (independent of the choice of $p \in L$). We define

$$\vartheta_{(F,k)}([(L, \nabla)]) = \sum_{\beta \in H_2(X, L, \mathbb{Z})} N_\beta z^\beta([(L, \nabla)]).$$

(Note that the sum may not converge in general.)

Example 4-5. Let $U = T = N \otimes \mathbb{C}^\times \simeq (\mathbb{C}^\times)^n$. Let $(\bar{X}, \bar{D})$ be a toric compactification and $F \subset \bar{D}$ a boundary divisor corresponding to a primitive vector $v \in N$. Let $g\colon \mathbb{C}^\times \to T$ be the associated 1-parameter subgroup of T. Then g extends to a morphism $\bar{g}\colon \mathbb{C} \to \bar{X}$ such that $\bar{g}$ meets F transversely at a single point and is disjoint from the other boundary divisors. Let h be the restriction of $\bar{g}$ to the closed unit disc $\mathbb{D} \subset \mathbb{C}$. Let $K \subset T \subset \bar{X}$ be the compact torus. Then $h\colon (\mathbb{D}, \partial\mathbb{D}) \to (\bar{X}, K)$ is a holomorphic disc ending on the fiber K of the moment map and passing through the point $e \in K$. It is the unique such disc and counts with multiplicity 1. The same applies to any choice of fiber and marked point (because they are permuted simply transitively by T). Similarly, there is a unique disc meeting F with contact order k given by the multiple cover $\tilde{h}(z) = h(z^k)$. See Cho and Oh [2006, Theorems 5.3 and 6.1].

The functions ϑ_q as defined above are discontinuous in general, because the counts of holomorphic discs N_β ending on an SYZ fiber $L = f^{-1}(b)$ vary discontinuously with b. This is due to the existence of SYZ fibers which bound holomorphic discs in U. Such fibers lie over (thickened) real codimension 1 walls in the base B. In more detail, suppose $\{L_t\}_{t\in[0,1]}$ are the SYZ fibers over a path crossing a wall in the base. If L_{t_0} bounds a holomorphic disc in U, then there may exist a family of holomorphic discs h_t in X ending on L_t for $t < t_0$, such that the limit of h_t as $t \to t_0$ is a *stable disc* given by the union of two discs ending on L_{t_0}, one of which is contained in U, and such that this stable disc does *not* deform to a holomorphic disc in X ending on L_t for $t > t_0$.

Example 4-6. (Auroux [2009], Example 3.1.2.) Let $(\bar{X}, \bar{D}) = (\mathbb{C}^2_{z_1,z_2}, (z_1z_2 = 0))$ with Kähler form $\frac{1}{2\pi}\frac{i}{2}(dz_1\wedge d\bar{z}_1 + dz_2\wedge d\bar{z}_2)$. Let $\pi : X \to \bar{X}$ be the blow up of $p = (1,0) \in \bar{X}$ with exceptional curve E, and $D \subset X$ the strict transform of $\bar{D}$. As in Example 4-4, we have a Kähler form ω on X and a map $\bar{f} : X \to \bar{B}$ which restricts to a Lagrangian torus fibration $f : U \to B$ over the interior B of $\bar{B}$. Moreover, over the complement of a small neighborhood N of $\bar{f}(E)$ the map $\bar{f}$ agrees with the moment map $\mu : \bar{X} \to \mathbb{R}^2_{\geq 0}$, $(z_1, z_2) \mapsto \frac{1}{2}(|z_1|^2, |z_2|^2)$. The map $\bar{f} : X \to \bar{B} \simeq \mathbb{R}^2_{\geq 0}$ is defined by $\frac{1}{2}|\pi^* z_1|^2$ and μ_{S^1}, the moment map for the S^1 action on (X, ω), normalized so that $\mu_{S^1} = 0$ on the strict transform of the z_1-axis. (Then on the singular fiber $\mu_{S^1} = \int_E \omega = \epsilon > 0$.)

There is a real codimension 1 wall H in the base B defined by $|\pi^* z_1| = 1$. Note that $(\pi^* z_1 = 1) \subset U$ is the union of two copies of $\mathbb{A}^1$ meeting in a node: $E \setminus E \cap D$ and the strict transform of $(z_1 = 1)$. (The node is the singular point of the pinched torus fiber of f.) These curves are S^1-equivariant and map to the wall in the base with fibers the S^1-orbits (which collapse to the singular point at the pinched torus fiber). Thus each smooth fiber over the wall bounds a holomorphic disc in U contained in one of the two curves.

Now let $D_1 = (\pi^* z_1 = 0) \subset X$ and consider the associated function ϑ_{D_1} defined by counting holomorphic discs in X meeting D transversely at a point of D_1, ending on an SYZ fiber L, and passing through a marked point $p \in L$. We assume L lies over the region $R = B \setminus N$ where the fibration f agrees with the moment map μ. Note that $f^{-1}R \subset U \setminus E \simeq \bar{U}$. Let $\pi(p) = (\nu_1, \nu_2) \in \bar{U} = (\mathbb{C}^\times)^2_{z_1,z_2}$, so $L = (|\pi^* z_1| = |\nu_1|, |\pi^* z_2| = |\nu_2|)$. If $|\nu_1| < 1$ then there is a unique disc given by the strict transform of the disc $\mathbb{D} \to \bar{X} = \mathbb{C}^2$, $z \mapsto (\nu_1 z, \nu_2)$. If $|\nu_1| > 1$ there are two discs, one described as before and the second given by the strict transform of the disc $z \mapsto (\nu_1 z, \nu_2(\frac{\nu_1 z - 1}{\bar{\nu}_1 - z})/(\frac{\nu_1 - 1}{\bar{\nu}_1 - 1}))$. See Cho and Oh [2006, Theorem 5.3]. Writing $\beta_1, \beta_1' \in H_2(X, L)$ for the classes of the two discs, observe that $\beta_1' = \beta_1 + \alpha \in H_2(X, L)$ where α is the parallel transport of the class of the disc associated to the portion of the wall meeting R. (More precisely, if $\{L_t\}_{t\in[0,1]}$ are the fibers over a path γ in $R \subset B$ crossing the wall at time t_0 from $|\pi^* z_1| > 1$ to $|\pi^* z_1| < 1$, then the limit of the holomorphic disc in class β_1' ending on L_t as $t \to t_0$ from below is the union of the holomorphic discs ending on L_{t_0} in classes β_1 and α.) We thus have

$$\vartheta_{D_1} = \begin{cases} z^{\beta_1} & \text{if } |\pi^* z_1| < 1 \\ z^{\beta_1} + z^{\beta_1'} = z^{\beta_1} \cdot (1 + z^{\alpha}) & \text{if } |\pi^* z_1| > 1. \end{cases}$$

On the other hand, let D_2 be the strict transform of $(z_2 = 0)$. Then, with notation as above, there is a unique disc meeting D transversely at a point of D_2, ending on L, and passing through $p \in L$, given by the inverse image of the disc $z \mapsto (\nu_1, \nu_2 z)$ in $\bar{X} = \mathbb{C}^2$. (For $\nu_1 = 1$, this is the stable disc given by the union of the strict transform of disc in $\bar{X}$ (which is the disc associated to the wall) and the exceptional curve E.) Thus, writing $\beta_2 \in H_2(X, L)$ for the class of this disc, we have $\vartheta_{D_2} = z^{\beta_2}$.

We have defined (using the local holomorphic coordinates z^{γ}, $\gamma \in H_1(L, \mathbb{Z})$) a complex structure on the total space of the dual fibration $V^o \to B^o$ of the smooth locus of the SYZ fibration $f : U \to B$. However it is expected that this does *not* extend to a complex structure on a fibration $V \to B$. Roughly speaking, if $V \to B$ is a topological extension of the fibration $V^o \to B^o$, and $W \subset V$ is a neighborhood of a point $p \in V \setminus V^o$, there are too few holomorphic functions defined on $W \cap V^o$ for the complex structure to extend. For instance, some of the z^{γ} are not well defined due to the monodromy action on $H_1(L, \mathbb{Z})$. The naive definition of the mirror V^o must be corrected to account for discs ending on SYZ fibers. These glueing corrections are such that the ϑ_q define global holomorphic functions on V^o, and can be used to define an extension $V^o \subset V$.

For instance, in Example 4-6, the corrected mirror V^o is an analytic open subset of $\hat{V}^o = (\mathbb{C}^{\times})^2_{w_1,w_2} \cup (\mathbb{C}^{\times})^2_{w_1',w_2'}$, where the two torus charts correspond to the two connected components $|\pi^* z_1| < 1$ and $|\pi^* z_1| > 1$ of the complement of the wall H in the base B, the glueing is given by

$$(w_1, w_2) \mapsto (w_1(1 + c w_2)^{-1}, w_2),$$

and $w_1 = w_1' = z^{\beta_1}$ and $w_2 = w_2' = z^{\beta_2}$ on the naive mirror (and we have trivialized the local system $H_2(X, L)$ over the the region $R \subset B$ as above). The parameter c is given by $c = z^{-E}$, $z^E := \exp(-2\pi \int_E \omega)$, so that $c w_2 = z^{-E} z^{\beta_2} = z^{\alpha}$ (since $\beta_2 = \alpha + [E]$ in $H_2(X, L)$). Then ϑ_{D_1} and ϑ_{D_2} are the global functions on V^o which restrict to w_1 and w_2 in the first chart. In fact, defining $\hat{V} = \operatorname{Spec} H^0(\mathcal{O}_{\hat{V}^o})$, we have an isomorphism

$$\hat{V} \longrightarrow (uv = 1 + cw) \subset \mathbb{A}^2_{u,v} \times \mathbb{C}^{\times}_w$$

given by $u \mapsto w_1$, $v \mapsto w_1'^{-1}$, $w \mapsto w_2$, and $\hat{V}^o = \hat{V} \setminus \{q\}$ where $q \mapsto (0, 0, -1/c)$ (cf. Section 3.2). The mirror $V \subset \hat{V}$ equals $V^o \cup \{q\}$, an analytic open subset of the affine variety $\hat{V}$.

Remark 4-7. The point $q \in V$ should correspond under HMS to the pinched torus fiber of the SYZ fibration $f : U \to B$ regarded as an immersed Lagrangian S^2 (with the trivial $U(1)$ local system). See Seidel [2013, Lecture 11].

In general, the wall crossing transformations should take the following form, cf. Abouzaid, Auroux, and Katzarkov [2016], p. 207. Let $\{L_t\}_{t\in[0,1]}$ be the fibers of $f : U \to B$ over a path crossing a wall in the base B. Assume for simplicity that all the holomorphic discs in U bounded by the L_t have relative homology class some fixed $\alpha \in H_2(U, L)$. The boundaries of these discs sweep out a cycle $c \in H_{n-1}(L)$. Then the wall crossing transformation in the local coordinates z^γ, $\gamma \in H_1(L)$, is given by

$$z^\gamma \mapsto z^\gamma \cdot f(z^\alpha)^{c\cdot\gamma}$$

where $f(z^\alpha) = 1 + z^\alpha + \cdots \in \mathbb{Q}[[z^\alpha]]$ is a power series encoding virtual counts of multiple covers of the discs.

4.3 Symplectic cohomology. Suppose that U is a positive log Calabi–Yau variety with maximal boundary. Suppose V is HMS mirror to U, so that we have an equivalence $\mathcal{F}(U) \simeq D(V)$ between the wrapped Fukaya category of U and the derived category of coherent sheaves on V. Symplectic cohomology SH^* is a version of Hamiltonian Floer cohomology for noncompact symplectic manifolds Seidel [2008]. There is a *closed-open string map* $SH^*(U) \to HH^*(\mathcal{F}(U))$ which is conjectured to be an isomorphism, cf. Seidel [2002], §4. (Recently, Ganatra–Pardon–Shende and Chantraine–Dimitroglou-Rizell–Ghiggini–Golovko have announced results which, combined with Ganatra [2012], would establish this result.) Recall that

$$HH^n(D(V)) \simeq \oplus_{p+q=n} H^p(\wedge^q T_V)$$

(the Hochschild–Kostant–Rosenberg isomorphism), in particular,

$$HH^0(D(V)) \simeq H^0(\mathcal{O}_V)$$

Thus the above conjecture and HMS would yield an isomorphism of $\mathbb{C}$-algebras $SH^0(U) \simeq H^0(\mathcal{O}_V)$. In particular, assuming the mirror V is affine, it can be constructed as $V = \operatorname{Spec} SH^0(U)$.

Conjecturally, $SH^0(U)$ has a natural basis parametrized by $U^{\mathrm{trop}}(\mathbb{Z})$. (This was proved by Pascaleff in dimension 2 Pascaleff [2013]; there is ongoing work of Ganatra–Pomerleano on the general case.) We expect that this basis corresponds to the global functions ϑ_q, $q \in U^{\mathrm{trop}}(\mathbb{Z})$ under the above isomorphism $SH^0(U) \simeq H^0(\mathcal{O}_V)$ (where we define $\vartheta_0 = 1$). In particular, we expect that the ϑ_q, $q \in U^{\mathrm{trop}}(\mathbb{Z})$ form a basis of $H^0(\mathcal{O}_V)$.

4.4 The Fock–Goncharov mirror of a cluster variety. Fock and Goncharov [2006] defined a candidate for the mirror V of a cluster variety U by a simple combinatorial recipe which we reproduce in our notation here. We will give a partial justification for the Fock–Goncharov construction in Section 5.

Recall that $U = X \setminus D$ is described (up to codimension two) as a union of copies T_α, $\alpha \in A$ of the algebraic torus $T = N \otimes \mathbb{C}^\times \simeq (\mathbb{C}^\times)^n$ with transition maps given by compositions of mutations

$$\mu = \mu_{(m,v)} : T_\alpha \dashrightarrow T_\beta, \quad \mu^*(z^{m'}) = z^{m'} \cdot (1 + cz^m)^{-\langle m', v\rangle}$$

for some $c \in \mathbb{C}^\times$. In addition we have a non-degenerate log 2-form σ on U such that $\sigma|_{T_\alpha} = \bar{\sigma} \in \wedge^2 M_{\mathbb{C}}$ for each α. We assume that the sign of m above has been chosen according to the convention of §5.3.

The Fock–Goncharov mirror $(V, \sigma^\vee)$ is described as follows. Let $T^\vee = N^* \otimes \mathbb{C}^\times$ be the dual algebraic torus to $T = N \otimes \mathbb{C}^\times$. We write $N^\vee = H_1(T^\vee, \mathbb{Z}) = N^* = M$ and $M^\vee = (N^\vee)^* = N$. Then (up to codimension two) V is a union of copies $T_\alpha^\vee$, $\alpha \in A$ of $T^\vee$, with transition maps given by

$$\mu^\vee = \mu_{(v,-m)} : T_\alpha^\vee \dashrightarrow T_\beta^\vee \quad \mu^{\vee *}(z^{v'}) = z^{v'} \cdot (1 + c^\vee z^v)^{\langle v', m\rangle}$$

for some $c^\vee \in \mathbb{C}^\times$. Let

$$\phi : N_{\mathbb{C}} \to M_{\mathbb{C}}, \quad \phi(v) = \bar{\sigma}(v, \cdot)$$

be the isomorphism determined by the non-degenerate form $\bar{\sigma}$ on $N_{\mathbb{C}}$ and $\bar{\sigma}^\vee$ the form on $N_{\mathbb{C}}^\vee = M_{\mathbb{C}}$ given by

$$\bar{\sigma}^\vee(m_1, m_2) = \bar{\sigma}(\phi^{-1}(m_1), \phi^{-1}(m_2)).$$

Then the log 2-form $\sigma^\vee$ on V is given by $\sigma^\vee|_{T_\alpha^\vee} = \bar{\sigma}^\vee \in \wedge^2 M_{\mathbb{C}}^\vee$.

Equivalently, given the data $N, \bar{\sigma} \in \wedge^2 M_{\mathbb{C}}, m_i \in M, v_i \in N, \lambda_i \in \mathbb{C}^\times, i = 1, \ldots, r$ of Section 3.3 determining the cluster variety $U = X \setminus D$ in terms of a toric model $\pi : (X, D) \to (\bar{X}, \bar{D})$, the Fock–Goncharov mirror is associated to the data $N^\vee = M$, $\bar{\sigma}^\vee$, $v_i \in M^\vee$, $-m_i \in N^\vee$, and some $\lambda_i^\vee \in \mathbb{C}^\times$, $i = 1, \ldots, r$.

Remark 4-8. Recall that, for mirror Calabi–Yau varieties U and V, symplectic deformations of U correspond to complex deformations of V, and vice versa. In particular, if we regard U as a symplectic manifold (forgetting the complex structure) and V as a complex manifold (forgetting the Kähler form), then the parameters $\lambda_i^\vee$ for V are determined by the class of the symplectic form on U (and the parameters λ_i for U are irrelevant).

Remark 4-9. The Fock–Goncharov mirror construction is an involution. The isomorphism between U and the mirror of the mirror of U is given in the torus charts by the map $T \to T$, $t \mapsto t^{-1}$.

Remark 4-10. We expect that the mirror of a log Calabi–Yau variety U with maximal boundary is of the same type if and only if U is positive. If U is a positive cluster variety, we expect that the Fock–Goncharov mirror is the mirror in the sense of SYZ and HMS. For a general cluster variety U, we expect that the true mirror is an analytic open subset of the Fock–Goncharov mirror. Cf. the discussion of completion of the mirror via symplectic inflation in the positive case in Auroux [2009, Sec. 2.2].

Example 4-11. Let $\bar{X}$ be the smooth projective toric surface given by the complete fan in $\mathbb{R}^2$ with rays generated by $(1,0)$, $(0,1)$, $(-1,2)$, $(-1,1)$, $(-1,0)$, $(-1,-1)$, $(0,-1)$, $(1,-1)$, $(2,-1)$. The toric boundary $\bar{D} \subset \bar{X}$ is a cycle of smooth rational curves with self-intersection numbers $-2,-2,-1,-2,-2,-1,-2,-2,-1$. Let $\pi : X \to \bar{X}$ be the blow up of three points in the smooth locus of $\bar{D}$, one point on each of the (-1)-curves, and D the strict transform of $\bar{D}$. Then $U = X \setminus D$ is a cluster variety. The divisor $D = \sum D_i$ is a cycle of nine (-2)-curves; in particular the intersection matrix $(D_i \cdot D_j)$ is negative semi-definite, and U is *not* positive. It is expected (cf. Auroux, Katzarkov, and Orlov [2006], Auroux [2009], §5) that the mirror of U is the log Calabi–Yau surface $V = Y \setminus E$ where $Y = \mathbb{P}^2$ and $E \subset Y$ is a smooth elliptic curve. In particular, there does not exist an open inclusion of an algebraic torus $(\mathbb{C}^\times)^2$ in V, so V is not a cluster variety.

Remark 4-12. In dimension 2, we may assume (multiplying σ by a non-zero scalar) that $\mathbb{Z} \cdot \bar{\sigma} = \wedge^2 M$. Let $\psi : N \to M$ be the isomorphism given by $\psi(v) = -\bar{\sigma}(v, \cdot)$. Then $\psi(v_i) = -m_i$ and $\psi^*(v_i) = -\psi(v_i) = m_i$. So if we take $\lambda_i^\vee = \lambda_i$ then the isomorphism $\psi \otimes \mathbb{C}^\times : T \to T^\vee$ extends to an isomorphism $U \to V$. That is, in dimension two the Fock–Goncharov mirror V of U is deformation equivalent to U.

Note that 2-torus fibrations are self-dual by Poincaré duality, so SYZ mirrors are diffeomorphic in dimension 2. For U a log Calabi–Yau surface with maximal boundary, the Fock–Goncharov mirror construction is valid if and only if U is positive (cf. Gross, Hacking, and Keel [2015b], Keating [2015]), and in that case the mirror V is deformation equivalent to U.

5 Scattering diagrams

Given a cluster variety U together with a choice of toric model, we explain how to build a *scattering diagram* in $U^{\mathrm{trop}}(\mathbb{R})$. Heuristically, this is the tropicalization of the collection of walls in the base of the SYZ fibration together with the attached generating functions encoding counts of holomorphic discs in U ending on SYZ fibers described in Section 4.1. We use the scattering diagram to construct a canonical topological basis ϑ_q, $q \in U^{\mathrm{trop}}(\mathbb{Z})$ of the algebra of global functions on a formal completion of the Fock–Goncharov mirror family. We expect that when U is positive this basis is algebraic and defines a canonical basis of global functions on the Fock–Goncharov mirror. We prove this under certain hypotheses on U related to positivity.

5.1 Definitions and algorithmic construction of scattering diagrams. Let $A = \mathbb{C}[t_1, \dots, t_r]$ and $\mathfrak{m} = (t_1, \dots, t_r) \subset A$. We write $\hat{M}$ for the $\mathfrak{m}$-adic completion $\varprojlim M/\mathfrak{m}^l M$ of an A-module M.

Let $N \simeq \mathbb{Z}^n$ be a free abelian group of rank n, and write $M = N^*$. Let $\bar{\sigma} \in \wedge^2 M_{\mathbb{C}}$ be a non-degenerate skew form.

Definition 5-1. A *wall* is a pair $(\mathfrak{d}, f)$ consisting of a codimension 1 rational polyhedral cone $\mathfrak{d} \subset N_{\mathbb{R}}$ together with an attached function $f \in \widehat{A[N]}$ satisfying the following properties. Let $m \in M$ be a primitive vector (determined up to sign) such that $\mathfrak{d} \subset m^{\perp}$. Then there exists a primitive vector $v \in N$ such that

1. $\bar{\sigma}(v, \cdot) = \nu \cdot m$ for some $\nu \in \mathbb{C}^{\times}$,
2. $f \in \widehat{A[z^v]} \subset \widehat{A[N]}$, and
3. $f \equiv 1 \bmod \mathfrak{m} z^v \widehat{A[z^v]}$.

(Note in particular $v \in m^{\perp}$ because $\bar{\sigma}$ is skew-symmetric.)

The cone $\mathfrak{d}$ is called the *support* of the wall. The vector $-v$ is called the *direction* of the wall. We say a wall $(\mathfrak{d}, f)$ is *incoming* if $v \in \mathfrak{d}$, otherwise, we say it is *outgoing*. (The terminology comes from the dimension 2 case, where the support of an outgoing wall is necessarily the ray $\mathbb{R}_{\geq 0} \cdot (-v)$ in the direction of the wall.)

Crossing a wall $(\mathfrak{d}, f)$ defines an associated automorphism θ of $\widehat{A[N]}$ over $\hat{A}$ such that $\theta \equiv \mathrm{id} \bmod \mathfrak{m}$. Let $m \in M$ be as in Definition 5-1. Then the automorphism associated to crossing the wall from $(m > 0)$ to $(m < 0)$ is given by

$$\theta \colon \widehat{A[N]} \to \widehat{A[N]}, \quad z^u \mapsto z^u \cdot f^{\langle u, m \rangle}.$$

A *scattering diagram* $\mathfrak{D}$ is a collection of walls such that for all $l \in \mathbb{N}$, there are finitely many walls $(\mathfrak{d}, f)$ such that $f \not\equiv 1 \bmod \mathfrak{m}^l$ (so that the associated automorphism is non-trivial modulo $\mathfrak{m}^l$).

The *support* $\operatorname{Supp} \mathfrak{D}$ of $\mathfrak{D}$ is the union of the supports of the walls. A *joint* of $\mathfrak{D}$ is an intersection of walls of codimension 2 in $N_{\mathbb{R}}$. The *singular locus* $\operatorname{Sing} \mathfrak{D}$ is the union of the joints of $\mathfrak{D}$ and the relative boundaries of the walls of $\mathfrak{D}$. A *chamber* of $\mathfrak{D}$ is the closure of an open connected component of $N_{\mathbb{R}} \setminus \operatorname{Supp} \mathfrak{D}$.

If $\gamma \colon [0, 1] \to N_{\mathbb{R}} \setminus \operatorname{Sing} \mathfrak{D}$ is a smooth path such that $\gamma(0), \gamma(1) \notin \operatorname{Supp}(\mathfrak{D})$ and γ is transverse to each wall it crosses, it defines an automorphism $\theta_{\mathfrak{D},\gamma}$ given by composing wall crossing automorphisms. In more detail, let $\mathfrak{D}_l \subset \mathfrak{D}$ be the finite subset of the scattering diagram consisting of walls $(\mathfrak{d}, f)$ such that $f \not\equiv 1 \bmod \mathfrak{m}^l$. Let $0 < t_1 < \dots < t_k < 1$ be the times at which γ crosses a wall of $\mathfrak{D}_l$, and θ_i the composition of the automorphisms associated to the walls crossed at time t_i (note that if two walls lie in the same hyperplane then the associated automorphisms commute, so θ_i is well defined). Let $\theta^l_{\mathfrak{D},\gamma}$ be the automorphism $\theta_k \circ \cdots \circ \theta_1$ of $(A/\mathfrak{m}^l)[N]$. Then $\theta_{\mathfrak{D},\gamma} = \varprojlim \theta^l_{\mathfrak{D},\gamma}$.

We say two scattering diagrams $\mathfrak{D}$, $\mathfrak{D}'$ are *equivalent* if $\theta_{\mathfrak{D},\gamma} = \theta_{\mathfrak{D}',\gamma}$ for all paths γ such that $\theta_{\mathfrak{D},\gamma}$ and $\theta_{\mathfrak{D}',\gamma}$ are defined.

A version of the following result was proved in dimension two in Kontsevich and Soibelman [2006], §10. The general case follows from Gross and Siebert [2011].

Theorem 5-2. *(Gross, Hacking, Keel, and Kontsevich [2018], Theorem 1.12) Let $\mathfrak{D}_{\text{in}}$ be a scattering diagram such that the support of each wall is a hyperplane. Then there is a scattering diagram $\mathfrak{D} = \operatorname{Scatter}(\mathfrak{D}_{\text{in}})$ containing $\mathfrak{D}_{\text{in}}$ such that*

1. *$\mathfrak{D} \setminus \mathfrak{D}_{\text{in}}$ consists of outgoing walls, and*
2. *$\theta_{\mathfrak{D},\gamma} = \mathrm{id}$ for all loops γ such that $\theta_{\mathfrak{D},\gamma}$ is defined.*

Moreover, $\mathfrak{D}$ is uniquely determined up to equivalence by these properties.

The theorem is proved modulo $\mathfrak{m}^l$ for each $l \in \mathbb{N}$ by induction on l. The inductive step is an explicit algorithmic construction. A self-contained proof in dimension two is given in Gross, Pandharipande, and Siebert [2010], Theorem 1.4. The basic construction in the general case is the same, cf. Gross, Hacking, Keel, and Kontsevich [2018], Appendix C.

5.2 Initial scattering diagram for cluster variety. Let (U, σ) be a cluster variety. Recall the combinatorial data from Section 3.3 describing U in terms of a toric model $\pi\colon (X, D) \to (\bar{X}, \bar{D})$: Let $T = \bar{X} \setminus \bar{D} \simeq (\mathbb{C}^\times)^n$ be the big torus, $N = H_1(T, \mathbb{Z}) \simeq \mathbb{Z}^n$, and $M = N^*$. We have $\bar{\sigma} = \sigma|_T \in H^0(\Omega^2_{\bar{X}}(\log \bar{D})) = \wedge^2 M_{\mathbb{C}}$ a non-degenerate skew matrix. We have primitive vectors $m_i \in M$, $v_i \in N$, $i = 1, \ldots, r$ such that $\bar{\sigma}(v_i, \cdot) = \nu_i m_i$, some $\nu_i \in \mathbb{C}^\times$. The rays $\mathbb{R}_{\geq 0} \cdot v_i$ are contained in the fan of $\bar{X}$ so correspond to components $\bar{D}_i \subset \bar{D}$. Then π is given by the blow up of the smooth centers

$$Z_i = \bar{D}_i \cap \overline{(z^{m_i} = \lambda_i)} \subset \bar{X}$$

for some $\lambda_i \in \mathbb{C}^\times$.

For the cluster variety U, we define

$$\mathfrak{D}_{\text{in}} = \{(m_i^\perp, 1 + t_i z^{v_i}) \mid i = 1, \ldots, r\}.$$

The enumerative interpretation is as follows. The strict transform of the divisor $\overline{(z^{m_i} = \lambda_i)} \subset \bar{X}$ in U is swept out by holomorphic discs ending on SYZ fibers L with boundary class $v_i \in H_1(L, \mathbb{Z}) = N$. These are the holomorphic discs corresponding to the ith initial wall. The two dimensional case is explained in Example 4-6. In dimension $n > 2$, U is locally isomorphic to a product $U' \times (\mathbb{C}^\times)^{n-2}$ (see Remark 3-5).

Example 5-3. Let $r = 2$ and $v_1, v_2 = (1,0), (0,1) \in N = \mathbb{Z}^2$. Then $\mathfrak{D} = \mathrm{Scatter}(\mathfrak{D}_{\mathrm{in}})$ consists of the two incoming walls $(\mathbb{R}\cdot(1,0), 1+t_1 z^{(1,0)})$ and $(\mathbb{R}\cdot(0,1), 1+t_2 z^{(0,1)})$ and one outgoing wall $(\mathbb{R}_{\geq 0}(-1,-1), 1+t_1t_2z^{(1,1)})$.

Here is the enumerative interpretation of the outgoing wall. The cluster variety U has toric model $\pi\colon (X,D) \to (\bar{X},\bar{D})$ where $\bar{X} = \mathbb{P}^2$ with toric boundary $\bar{D} = \bar{D}_1+\bar{D}_2+\bar{D}_3$, and π is given by blowing up two points p_1, p_2 in the smooth locus of $\bar{D}$, with $p_1 \in \bar{D}_1$ and $p_2 \in \bar{D}_2$. Let C be the strict transform of the line through p_1 and p_2. Then C meets D in a single point p. Holomorphic discs associated to the outgoing wall are approximated by holomorphic discs contained in $C \setminus \{p\}$. (One can also give an explicit description using Cho and Oh [2006] as in Example 4-6.)

Note that in general the walls of the scattering diagram may be dense in some regions of $N_{\mathbb{R}}$, and the attached functions are not polynomial. See e.g. Gross, Pandharipande, and Siebert [2010], Example 1.6 and Remark 5-5 below.

For $U = X \setminus D$ a log Calabi–Yau surface with maximal boundary, Gross, Pandharipande, and Siebert [ibid.] proves an enumerative interpretation of the scattering diagram in terms of virtual counts of maps $f\colon \mathbb{P}^1 \to X$ meeting the boundary D in a single point. A similar interpretation in terms of log Gromov-Witten invariants is expected in general Gross and Siebert [2016], §2.4.

The following lemma (which will be needed in Section 5.4 below) is left as an exercise.

Lemma 5-4. *Let U be a cluster variety with associated combinatorial data $v_i \in N$, $m_i \in M$, $i = 1,\dots,r$. Then*

$$\mathrm{Pic}(U) = \mathrm{im}(H^2(X,\mathbb{Z}) \to H^2(U,\mathbb{Z})) = \mathrm{coker}((v_1,\dots,v_r)^T : M \to \mathbb{Z}^r)$$

and $\pi_1(U) = H_1(U,\mathbb{Z}) = \mathrm{coker}((v_1,\dots,v_r)\colon \mathbb{Z}^r \to N)$.

5.3 Reduction to irreducible case and sign convention.

For (U,σ) a cluster variety such that $H_1(U,\mathbb{Q}) = 0$, there is an étale cover $\tilde{U} \to U$ which decomposes as a product of cluster varieties (U_i,σ_i), $U_i = X_i \setminus D_i$, such that $H^0(\Omega^2_{X_i}(\log D_i)) = \mathbb{C}\cdot\sigma_i$ for each i (cf. the Bogomolov decomposition theorem in the compact setting).

We now assume that $H^0(\Omega^2_X(\log D)) = \mathbb{C}\cdot\sigma$. It follows from Section 3.1 that the subspace $H^0(\Omega^2_X(\log D)) \subset H^0(\Omega^2_{\bar{X}}(\log \bar{D})) = \wedge^2 M_{\mathbb{C}}$ is defined over $\mathbb{Q}$. So we may assume, multiplying by a nonzero scalar, that $\sigma \in \wedge^2 M$. Then $\bar{\sigma}(v_i,\cdot) = \nu_i m_i$ for some $\nu_i \in \mathbb{Q}$. Note that the blow up description of $U = X \setminus D$ depends on m_i through the center $Z_i = D_i \cap \overline{(z^{m_i} = \lambda_i)}$. So we may assume (replacing m_i by $-m_i$ and λ_i by λ_i^{-1} if necessary) that $\nu_i > 0$.

In this case, it follows from the proof of Theorem 5-2 that for all walls $(\mathfrak{d}, f)$ in $\mathfrak{D} = \mathrm{Scatter}(\mathfrak{D}_{\mathrm{in}})$ we have $\mathfrak{d} \subset m^{\perp}$ for some nonzero $m \in M$ such that $m = \sum a_i m_i$ with

$a_i \geq 0$ for each i. In particular, if the m_i are linearly independent then $\mathfrak{D}$ has two chambers given by

$$\mathcal{C}^+ = \{v \in N_{\mathbb{R}} \mid \langle m_i, v\rangle \leq 0 \text{ for all } i = 1, \dots, r\}$$

and $\mathcal{C}^- := -\mathcal{C}^+$.

5.4 Reduction to the case of linearly independent m_i: Universal deformation and universal torsor. We now explain how to reduce to the case that the $m_i \in M$ are linearly independent. Assume for simplicity that M is generated by the m_i. Equivalently, by Lemma 5-4, the Fock–Goncharov mirror V of U is simply connected.

The surjection $(m_1, \dots, m_r)\colon \mathbb{Z}^r \to M$ determines a surjective homomorphism $\varphi\colon (\mathbb{C}^\times)^r \to T^\vee$ and, dually, an injective homomorphism $T \hookrightarrow (\mathbb{C}^\times)^r$

We have the universal deformation $p\colon \mathcal{U} \to S := (\mathbb{C}^\times)^r/T$ of $U = p^{-1}([(\lambda_i)])$ given by varying the parameters λ_i.

Our assumption implies that $\operatorname{Pic} V = \operatorname{coker}(m_1, \dots, m_r)^T$ is torsion-free by Lemma 5-4. Let $L_1, \dots, L_s$ be a basis of $\operatorname{Pic} V$. The *universal torsor* $q\colon \tilde{V} \to V$ is the fiber product of the $\mathbb{C}^\times$-bundles $L_i^\times$ over V. It is a principal bundle with group $\operatorname{Hom}(\operatorname{Pic} V, \mathbb{C}^\times) = \ker(\varphi)$.

The 2-form σ on U lifts canonically to a *relative* 2-form $\sigma_{\mathcal{U}}$ on $\mathcal{U}/S$ (non-degenerate on each fiber). Equivalently, $\sigma_{\mathcal{U}}$ defines a Poisson bracket on $\mathcal{U}$ with symplectic leaves the fibers of p. The 2-form $\sigma^\vee$ on V pulls back to a degenerate 2-form on $\tilde{V}$.

Write $N_{\mathcal{U}} = \mathbb{Z}^r$ with standard basis $e_1, \dots, e_r$ and $M_{\mathcal{U}} = N_{\mathcal{U}}^*$ with dual basis $f_1, \dots, f_r$. We have the inclusion $(m_1, \dots, m_r)^T\colon N \subset N_{\mathcal{U}}$. The variety $(\mathcal{U}, \sigma_{\mathcal{U}})$ is a (generalized) cluster variety with toric model given by the combinatorial data $N_{\mathcal{U}}$, $\bar{\sigma}_{\mathcal{U}} = \bar{\sigma} \in \wedge^2 M_{\mathbb{C}}$, $v_i \in N_{\mathcal{U}}$, $f_i \in M_{\mathcal{U}}$, $i = 1, \dots, r$. The variety $(\tilde{V}, q^*\sigma^\vee)$ is the (generalized) Fock–Goncharov mirror cluster variety. Cf. Gross, Hacking, and Keel [2015a], §4.

Roughly speaking, in the terminology of Fock and Goncharov, $\mathcal{U}$ is the $\mathcal{X}$-variety for the given combinatorial data and $\tilde{V}$ is the $\mathcal{A}$-variety for the Langlands dual data. The Fomin–Zelevinsky (upper) cluster algebra is the ring of global functions $H^0(\tilde{V}, \mathcal{O}_{\tilde{V}})$.

One can construct the scattering diagram associated to $\mathcal{U}$ in $(N_{\mathcal{U}})_{\mathbb{R}} \supset N_{\mathbb{R}}$ as before using the relative 2-form $\bar{\sigma}_{\mathcal{U}} = \bar{\sigma} \in \wedge^2 M_{\mathbb{C}}$. (Condition (1) in Definition 5-1 can be rewritten $\bar{\sigma}^\vee(-m, \cdot) = \nu^\vee v$, where $\nu^\vee = \nu^{-1}$, that is, the corresponding condition for the Fock–Goncharov mirror V. In this form it generalizes to the above setting.)

Note that $H^0(\tilde{V}, \mathcal{O}_{\tilde{V}}) = \oplus_{L \in \operatorname{Pic} V} H^0(V, L)$ is the Cox ring of V. The torus $\operatorname{Hom}(\operatorname{Pic} V, \mathbb{C}^\times)$ acts with weight $L \in \operatorname{Pic} V$ on the summand $H^0(V, L)$. Our construction of a canonical basis of $H^0(\tilde{V}, \mathcal{O}_{\tilde{V}})$ is equivariant for the torus action. In particular we obtain a canonical basis of $H^0(V, \mathcal{O}_V)$ (and also $H^0(V, L)$ for each $L \in \operatorname{Pic} V$). Thus we may replace (U, σ) and $(V, \sigma^\vee)$ by $(\mathfrak{U}, \sigma_{\mathfrak{U}})$ and $(\tilde{V}, q^*\sigma^\vee)$ and assume that the m_i are linearly independent.

5.5 Mutation invariance of the support of the scattering diagram and cluster complex. The support of the scattering diagram $\mathfrak{D}$ is invariant under mutation Gross, Hacking, Keel, and Kontsevich [2018, Sec. 1.3]. That is, if $\pi\colon (X, D) \to (\bar{X}, \bar{D})$ and $\pi'\colon (X, D) \to (\bar{X}', \bar{D}')$ are two toric models for (X, D) related by a mutation $\mu\colon T \dashrightarrow T'$ as in (3-7), and $\mathfrak{D}$, $\mathfrak{D}'$ are the scattering diagrams associated to π, π' then $\mu^{\text{trop}}(\operatorname{Supp} \mathfrak{D}) = \operatorname{Supp} \mathfrak{D}'$. Heuristically, this is so because $\operatorname{Supp} \mathfrak{D}$ is the union of the tropicalizations of all holomorphic discs in U ending on SYZ fibers, viewed in $N_{\mathbb{R}}$ using the $\mathbb{Z}$PL identification $U^{\text{trop}}(\mathbb{R}) \simeq N_{\mathbb{R}}$ corresponding to the open inclusion $T = \bar{X} \setminus \bar{D} \subset U = X \setminus D$ of log Calabi–Yau varieties.

Let $\mu = \mu_{(m,v)}$ as in (3-7). Let (m_i, v_i), $i = 1, \dots, r$, be the combinatorial data for the toric model π, with $(m, v) = (m_1, v_1)$. Recall the explicit formula in Example 3-10 for μ^{trop}. In particular, μ^{trop} is linear on the halfspaces $\mathcal{H}_+ = (m \geq 0)$ and $\mathcal{H}_- = (m \leq 0)$. Let T_+, T_- be the linear automorphisms of $N_{\mathbb{R}}$ which agree with μ^{trop} on $\mathcal{H}_+$ and $\mathcal{H}_-$. Then $T_+ = \text{id}$ and T_- is the symplectic transvection

$$T_-(w) = w - \langle m, w \rangle v = w - \nu^{-1} \cdot \bar{\sigma}(v, w) v.$$

Then the combinatorial data (m_i', v_i'), $i = 1, \dots, r$ for π' is given by

$$(m_i', v_i') = \begin{cases} (-m_1, -v_1) & \text{if } i = 1 \\ ((T_+^*)^{-1}(m_i), T_+(v_i)) & \text{if } v_i \in \mathcal{H}_+ \text{ and } i > 1 \\ ((T_-^*)^{-1}(m_i), T_-(v_i)) & \text{if } v_i \in \mathcal{H}_- \text{ and } i > 1. \end{cases}$$

(The sign reversal $v_1' = -v_1$ follows from the description of the elementary transformation in Section 3.2. The signs of the m_i' are determined by the sign convention of Section 5.3.)

Using the explicit formula in Example 3-10 for μ^{trop}, we see that the chambers $\mu^{\text{trop}}(\mathcal{C}^+)$ and $\mathcal{C}^{+'}$ in $\mathfrak{D}'$ meet along the codimension 1 face defined by $m = 0$.

Applying elementary transformations repeatedly, we obtain a simplicial fan $\Delta^+ \subset U^{\text{trop}}(\mathbb{R}) \simeq N_{\mathbb{R}}$ with maximal cones the positive chambers $\mathcal{C}_\alpha^+$ associated to each torus chart T_α, such that two maximal cones meet along a codimension 1 face if and only if the torus charts are related by a mutation. This is the Fock–Goncharov *cluster complex* (the dual graph is the Fomin–Zelevinsky *exchange graph*). The maximal cones of Δ^+ are chambers of the scattering diagram $\mathfrak{D}$. Thus the scattering diagram is discrete in the interior of the support of Δ^+ in $N_{\mathbb{R}}$.

Remark 5-5. Note that here we are using our assumption that the m_i are linearly independent (see Section 5.4). Without this assumption, the scattering diagram can be everywhere dense in $N_{\mathbb{R}}$. For example, this is the case for $U = X \setminus D$ where $X \subset \mathbb{P}^3$ is a smooth cubic surface and D is a triangle of lines on X (equivalently, (X, D) is obtained from $\bar{X} = \mathbb{P}^2$ together with its toric boundary $\bar{D}$ by blowing up six general points in the smooth locus of $\bar{D}$, two on each line). To see this, first observe that we can construct another toric model of U of the same combinatorial type as follows. Let $\bar{D}_1, \bar{D}_2, \bar{D}_3$ be the components of $\bar{D}$. Let $\bar{X}^1$ be the blowup of the point $p = \bar{D}_1 \cap \bar{D}_2 \in \bar{X}$. Then $\bar{X}^1$ is a ruled surface with sections the exceptional divisor E and the strict transform of $\bar{D}_3$. Let p_1, p_2 be the two centers of $\pi : X \to \bar{X}$ on $\bar{D}_3$ and let $\bar{X}^1 \dashrightarrow \bar{X}^2$ be the composite of the elementary transformations with centers p_1 and p_2. Finally, blow down the strict transform of $\bar{D}_3$ to obtain $\bar{X}' \simeq \mathbb{P}^2$ with toric boundary $\bar{D}'$ given by the strict transforms of $\bar{D}_1$, $\bar{D}_2$, and E. Then by construction we have another toric model $\pi' \colon (X', D') \to (\bar{X}', \bar{D}')$ of U given by blowing up two points on each boundary divisor. The rational map $T = \bar{X} \setminus \bar{D} \dashrightarrow T' = \bar{X}' \setminus \bar{D}'$ is a composite of two mutations. Let $v_1, v_2 \in N$ correspond to the boundary divisors $\bar{D}_1, \bar{D}_2$ of $\bar{X}$ under the identification $U^{\mathrm{trop}}(\mathbb{Z}) = N$ given by $T \subset U$, then the boundary divisor E of $\bar{X}'$ correponds to $v_1 + v_2 \in N$ under this identification. Recall that the scattering diagram associated to a toric model has an incoming wall associated to each blow up. The support of the wall contains the ray corresponding to the boundary divisor containing the center of the blow up. Moreover, the support of the scattering diagram is invariant under mutation. In particular, it follows that the rays generated by v_1, $v_1 + v_2$, and v_2 lie in $\mathrm{Supp}(\mathfrak{D})$. Repeating the above construction one can prove by induction that every rational ray lies in $\mathrm{Supp}(\mathfrak{D})$.

In terms of the construction of the versal deformation $\mathfrak{U}$ of U in Section 5.4, the scattering diagram $\mathfrak{D}$ for U is the slice of the scattering diagram $\mathfrak{D}_{\mathfrak{U}}$ for $\mathfrak{U}$ by the subspace $(m_1, \ldots, m_r) \colon N_{\mathbb{R}} \hookrightarrow \mathbb{R}^r$. This slice can miss the discrete part of $\mathfrak{D}_{\mathfrak{U}}$ so that there are no chambers in $\mathfrak{D}$.

The functions attached to walls of the scattering diagram change in a simple way under mutation. Each wall of the cluster complex corresponds to a portion of an incoming wall for some seed (and there are no outgoing walls with support contained in an incoming wall Gross, Hacking, Keel, and Kontsevich [2018], Remark 1.29). So one can describe the functions attached to walls of the cluster complex explicitly. One finds that they are polynomials (in fact of the form $1+cz^v$ for $c \in A = \mathbb{C}[t_1, \ldots, t_r]$ a monomial and $v \in N$). So one can define an algebraic family $\mathcal{V}/\mathbb{A}^r_{t_1,\ldots,t_r}$ as follows: $\mathcal{V}$ is a union of copies of $T^\vee \times \mathbb{A}^r$ indexed by chambers $\mathcal{C}^+ \in \Delta^+$, with transition maps $T^\vee \times \mathbb{A}^r \dashrightarrow T^\vee \times \mathbb{A}^r$ given by $\theta_{\mathfrak{D},\gamma}$ for γ a path in $\mathrm{Supp}\,\Delta^+$ from the first chamber to the second. One finds that the restriction of $\mathcal{V}/\mathbb{A}^r$ to $(\mathbb{C}^\times)^r$ is the family of Fock–Goncharov mirrors to U. Moreover, because the automorphisms are trivial modulo $\mathfrak{m} = (t_1, \ldots, t_r)$, the special fiber $\mathcal{V}_0$ equals the torus $T^\vee$.

5.6 Broken lines. We now describe the construction of global functions $\vartheta_q, q \in U^{\text{trop}}(\mathbb{Z})$ on the mirror family $\mathcal{V}/\mathbb{A}^r$ using a tropical analogue of the heuristic construction described in Section 4.2.

Definition 5-6. Let $v \in U^{\text{trop}}(\mathbb{Z}) = N$ be a nonzero vector and $p \in N_{\mathbb{R}}$ a general point. A *broken line* for v with endpoint p is a continuous piecewise-linear path $\gamma : (-\infty, 0] \to N_{\mathbb{R}}$ together with, for each domain of linearity $L \subset (-\infty, 0]$, a monomial $c_L \cdot z^{v_L}$, $c_L \in A = \mathbb{C}[t_1, \dots, t_r]$, $v_L \in N$, such that

1. There is an initial unbounded domain of linearity with attached monomial $1 \cdot z^v$.
2. For all L and $t \in L$, $\gamma'(t) = -v_L$.
3. If γ is not linear at $t \in (-\infty, 0]$ then γ crosses a wall at time t. Let L and L' be the domains of linearity before and after crossing the wall and θ the wall crossing automorphism. Then $c_{L'} z^{v_{L'}}$ is a monomial term in $\theta(c_L z^{v_L}) \in \widehat{A[N]}$.
4. $\gamma(0) = p$.

We write $M(\gamma)$ for the final monomial attached to a broken line γ.

Recall that the Fock–Goncharov mirror V of U is a union $V = \bigcup_{\mathcal{C}^+ \in \Delta^+} T^\vee_{\mathcal{C}^+}$ of copies of the dual torus $T^\vee = M \otimes \mathbb{C}^\times$ indexed by the maximal cones $\mathcal{C}^+$ of the cluster complex Δ^+. We have the family $\mathcal{V} \to \operatorname{Spec} A = \mathbb{A}^r_{t_1,\dots,t_r}$, $\mathcal{V} = \bigcup_{\mathcal{C}^+ \in \Delta^+} T^\vee_{\mathcal{C}^+} \times \mathbb{A}^r$, with fiber V over the point $t_i = c_i^\vee = -1/\lambda_i^\vee$, and its formal completion $\hat{\mathcal{V}} \to \operatorname{Spf} \hat{A}$ over $0 \in \mathbb{A}^r$.

We now define theta functions ϑ_v for $v \in U^{\text{trop}}(\mathbb{Z}) \simeq N$ on $\hat{\mathcal{V}}$. We define $\vartheta_0 = 1$. Let $v \in N$ be a nonzero vector. For $p \in \operatorname{Supp} \Delta^+$ a general point, we define

$$\vartheta_{v,p} = \sum_\gamma M(\gamma) \in \widehat{A[N]},$$

where the sum is over broken lines γ for v with endpoint p. For general points $p, p' \in \operatorname{Supp} \Delta^+$, and γ a path from p to p', we have $\theta_{\mathfrak{D},\gamma}(\vartheta_{v,p}) = \vartheta_{v,p'}$ Gross, Hacking, Keel, and Kontsevich [ibid.], Theorem 3.5, Carl, Pumperla, and Siebert [2010], §4. So, by the definition of $\mathcal{V}/\mathbb{A}^r$, the $\vartheta_{v,p}$ for $p \in \operatorname{Supp} \Delta^+$ general define a global function ϑ_v on $\hat{\mathcal{V}}$.

Example 5-7. Consider the scattering diagram $\mathfrak{D}$ for the data $r = 1$, $v_1 = (1,0) \in N = \mathbb{Z}^2$. Write $z_1 = z^{(1,0)}$ and $z_2 = z^{(0,1)}$. The scattering diagram $\mathfrak{D} = \mathfrak{D}_{\text{in}}$ consists of the single wall $(\mathbb{R} \cdot (1,0), 1 + tz_1)$. Let $v = (0,1)$. For $p = (a,b) \in N_{\mathbb{R}} = \mathbb{R}^2$, if $b > 0$ there is a unique broken line for v with endpoint p given by $\gamma(t) = (a,b) - t(0,1)$ with attached monomial z_2. If $b < 0$ there are two broken lines, one described as before and the second given by $\gamma(t) = (a-b, b) - t(0,1)$ for $t \leq b$ with attached monomial z_2 and

$\gamma(t) = (a,b) - t(1,1)$ for $b \le t \le 0$, with attached monomial tz_1z_2. Thus $\vartheta_{v,p} = z_2$ for $b > 0$ and $\vartheta_{v,p} = z_2 + tz_1z_2 = z_2(1+tz_1)$ for $b < 0$. This is the tropical version of Example 4-6. See Cheung, Gross, Muller, Musiker, Rupel, Stella, and Williams [2017] for more examples.

Recall that $\mathcal{V}_0 = T^\vee$, and note that ϑ_v restricts to the character z^v on $\mathcal{V}_0$ (because $M(\gamma) \equiv 0 \bmod \mathfrak{m}$ for any broken line that bends). So the ϑ_v, $v \in U^{\text{trop}}(\mathbb{Z})$ restrict to a basis of $H^0(\mathcal{V}_0, \mathcal{O}_{\mathcal{V}_0})$. It follows that the ϑ_v, $v \in U^{\text{trop}}(\mathbb{Z})$ define a topological basis of $H^0(\hat{\mathcal{V}}, \mathcal{O}_{\hat{\mathcal{V}}})$. That is, for every element $f \in H^0(\hat{\mathcal{V}}, \mathcal{O}_{\hat{\mathcal{V}}})$ there is a unique expression $f = \sum_{v \in U^{\text{trop}}(\mathbb{Z})} a_v \vartheta_v$ where $a_v \in \hat{A}$ for each v and for all $l \in \mathbb{N}$ there are finitely many a_v such that $a_v \not\equiv 0 \bmod \mathfrak{m}^l$.

The formal function ϑ_v defines a function on $\mathcal{V}$ (and so on the fiber V) if and only if the local expressions $\vartheta_{v,p} \in \widehat{A[N]}$ lie in $A[N]$, that is, are Laurent polynomials with coefficients in A. This is not the case in general. However, one can show that if $\vartheta_{v,p}$ is a Laurent polynomial for some $p \in \operatorname{Supp} \Delta^+$, then the same is true for all p, so that ϑ_v lies in $H^0(\mathcal{V}, \mathcal{O}_{\mathcal{V}})$ Gross, Hacking, Keel, and Kontsevich [2018], Proposition 7.1.

Example 5-8. Let $\mathcal{C}^+$ be a chamber of Δ^+ and $v \in \mathcal{C}^+ \cap U^{\text{trop}}(\mathbb{Z})$ an integral point. Let $p \in \mathcal{C}^+$ be a general point. Then there is a unique broken line for v and p, given by $\gamma\colon (-\infty, 0] \to N_{\mathbb{R}}$, $\gamma(t) = p - tv$, with attached monomial z^v. See Gross, Hacking, Keel, and Kontsevich [ibid., Corollary 3.9]. It follows that ϑ_v is a global function on V such that $\vartheta_v|_{T^\vee_{\mathcal{C}^+}} = z^v$. In the terminology of Fomin–Zelevinsky, ϑ_v is a *cluster monomial*.

Example 5-9. For cluster algebras of *finite type* Fomin and Zelevinsky [2003], the cluster complex has finitely many cones and is a complete fan, that is, $\operatorname{Supp} \Delta^+ = U^{\text{trop}}(\mathbb{R}) \simeq N_{\mathbb{R}}$. In this case, every theta function is a cluster monomial, and the cluster monomials form a basis of $H^0(V, \mathcal{O}_V)$.

Cluster algebras of finite type correspond to finite root systems Fomin and Zelevinsky [ibid.]. The mirror U of the A_2-cluster variety V is described in Example 5-3. Let $\vartheta_1, \ldots, \vartheta_5$ be the theta functions on $\mathcal{V}/\mathbb{A}^2$ corresponding to the primitive generators of the rays of the cluster complex in cyclic order. We identify V with the fiber of $\mathcal{V}$ over $(1,1) \in \mathbb{A}^2$. The theta function basis of $H^0(V, \mathcal{O}_V)$ is given by the cluster monomials

$$\{\vartheta_i^a \vartheta_{i+1}^b \mid a, b \in \mathbb{Z}_{\ge 0},\ i \in \mathbb{Z}/5\mathbb{Z}\}.$$

The algebra structure is given by

$$V = (\vartheta_{i-1}\vartheta_{i+1} = \vartheta_i + 1,\ i \in \mathbb{Z}/5\mathbb{Z}) \subset \mathbb{A}^5_{\vartheta_1,\ldots,\vartheta_5}.$$

The closure of V in $\mathbb{P}^5$ is the del Pezzo surface of degree 5 (the blowup of 4 points in $\mathbb{P}^2$ in general position) with an anti-canonical cycle of 5 (-1)-curves at infinity. In this case, the mirror V of U is isomorphic to U (since U is rigid this is a special case of Remark 4-12).

Theorem 5-10. *(Gross, Hacking, Keel, and Kontsevich [2018], Proposition 0.7) Let $U^{\text{trop}}(\mathbb{R}) = N_{\mathbb{R}}$ be the identification associated to some toric model of U. Suppose that the support of the cluster complex Δ^+ is not contained in a half-space in $N_{\mathbb{R}}$ under this identification. Then each ϑ_v defines a global function on $\mathcal{V}$, and the ϑ_v, $v \in U^{\text{trop}}(\mathbb{Z})$ define a $\mathbb{C}$-basis of $H^0(V, \mathcal{O}_V)$.*

See Gross, Hacking, Keel, and Kontsevich [ibid., Sec. 8], for the relation between the hypothesis and positivity of U.

Example 5-11. (Gross, Hacking, Keel, and Kontsevich [ibid., Corollary 0.20], Magee [2017], cf. Goncharov and Shen [2015]). Let $G = SL_m$. Let $B \subset G$ be a Borel subgroup, $N \subset B$ the maximal unipotent subgroup, and $H \subset B$ a maximal torus. Let $F = G/B$ be the full flag variety, and $\tilde{F} = G/N$ its universal torsor, a principal $H = B/N$-bundle over F. The variety F is called the *base affine space*. By the Borel–Weil–Bott theorem,

$$H^0(\tilde{F}, \mathcal{O}_{\tilde{F}}) = \oplus_{L \in \operatorname{Pic} F} H^0(F, L) = \oplus_\lambda V_\lambda,$$

the direct sum of the irreducible representations of G (where $\lambda \in \operatorname{Lie}(H)^*$ denotes a dominant weight). Cf. Fulton and Harris [1991], p. 392–3.

Let $B_- \subset G$ be the opposite Borel subgroup such that $B \cap B_- = H$. Let $V \subset F$ be the open subset of flags transverse to the flags with stabilizers B and B_-. Let $\tilde{V} \subset \tilde{F}$ be its inverse image. Then $\tilde{V}$ is identified with the double Bruhat cell $G^{w_0,e} := Bw_0B \cap B^- \subset G$ where $w_0 \in W = S_m$, $w_0(i) = m + 1 - i$, is the longest element of the Weyl group. In particular, $\tilde{V}$ is a cluster variety in the sense of Fomin–Zelevinsky Berenstein, Fomin, and Zelevinsky [2005]. (From our point of view, there are algebraic tori T_1 and T_2, an action of T_1 on $\tilde{V}$ and a T_1 invariant fibration $\tilde{V} \to T_2$, such that the quotients $\tilde{V}_t/T_1$ of the fibers are cluster varieties in the sense of Definition 3-1. Thus $\tilde{V}$ is given by a combination of the two constructions in Section 5.4.) We remark that the cluster structure on $\tilde{V}$ is closely related to the Poisson structure on G associated to the choice $H \subset B \subset G$ Gekhtman, Shapiro, and Vainshtein [2010, Sec. 1.3], which is the first order term of the non-commutative deformation of G to the quantum group Chari and Pressley [1995],§7.3.

One can show using a generalization of Theorem 5-10 that the theta functions give a canonical basis of $H^0(\tilde{V}, \mathcal{O}_{\tilde{V}})$. The variety $\tilde{F}$ is a partial compactification of $\tilde{V}$, such that $\tilde{F} \setminus \tilde{V}$ is a union of $2(m-1)$ boundary divisors along which the holomorphic volume form Ω on $\tilde{V}$ has a pole. Using positivity properties of the theta function basis $\mathcal{B}$ of $H^0(\tilde{V}, \mathcal{O}_{\tilde{V}})$, one can further show that the subalgebra $H^0(\tilde{F}, \mathcal{O}_{\tilde{F}})$ has basis given by the subset of $\mathcal{B}$ consisting of theta functions which are regular on $\tilde{F}$. The set of such functions is indexed by the integral points of a polyhedral cone in $\tilde{U}^{\text{trop}}(\mathbb{R})$ (where $\tilde{U}$ denotes the Fock–Goncharov mirror of $\tilde{V}$). This polyhedral cone C is given by $C = (W^{\text{trop}} \geq 0)$ where $W : \tilde{U} \to \mathbb{A}^1$ is the regular function given by the sum of the theta functions on

$\tilde{U}$ corresponding to the boundary divisors of the partial compactification $\tilde{V} \subset \tilde{F}$. (Thus, according to general principles, $\tilde{F}$ is mirror to the *Landau–Ginzburg model* $W : \tilde{U} \to \mathbb{A}^1$, cf. Abouzaid, Auroux, and Katzarkov [2016], §2.2.) For a specific choice of toric model of $\tilde{U}$, the cone C is identified with the Gel'fand–Tsetlin cone Goncharov and Shen [2015].

The theta function basis is equivariant for the action of H on $\tilde{F}$. Thus the theta functions with weight λ give a canonical basis of the irreducible representation V_λ.

Remark 5-12. The construction of theta functions given here appears to depend on the choice of a toric model of U. However, one can show, using the behavior of the scattering diagram under mutation, that the families $\mathcal{V}/\mathbb{A}^r$ together with the theta functions for different toric models are compatible Gross, Hacking, Keel, and Kontsevich [2018], Theorem 6.8. In the terminology of mirror symmetry, they correspond to different large complex structure limits of the mirror family.

References

Mohammed Abouzaid, Denis Auroux, and Ludmil Katzarkov (2016). "Lagrangian fibrations on blowups of toric varieties and mirror symmetry for hypersurfaces". *Publ. Math. Inst. Hautes Études Sci.* 123, pp. 199–282. MR: 3502098 (cit. on pp. 697, 698, 702, 714).

Dan Abramovich, Kalle Karu, Kenji Matsuki, and Jarosław Włodarczyk (2002). "Torification and factorization of birational maps". *J. Amer. Math. Soc.* 15.3, pp. 531–572. MR: 1896232 (cit. on p. 695).

Denis Auroux (2009). "Special Lagrangian fibrations, wall-crossing, and mirror symmetry". In: *Surveys in differential geometry. Vol. XIII. Geometry, analysis, and algebraic geometry: forty years of the Journal of Differential Geometry*. Vol. 13. Surv. Differ. Geom. Int. Press, Somerville, MA, pp. 1–47. MR: 2537081 (cit. on pp. 689, 699, 700, 704).

– (2014). "A beginner's introduction to Fukaya categories". In: *Contact and symplectic topology*. Vol. 26. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, pp. 85–136. MR: 3220941 (cit. on p. 698).

Denis Auroux, Ludmil Katzarkov, and Dmitri Orlov (2006). "Mirror symmetry for del Pezzo surfaces: vanishing cycles and coherent sheaves". *Invent. Math.* 166.3, pp. 537–582. MR: 2257391 (cit. on p. 704).

Arkady Berenstein, Sergey Fomin, and Andrei Zelevinsky (2005). "Cluster algebras. III. Upper bounds and double Bruhat cells". *Duke Math. J.* 126.1, pp. 1–52. MR: 2110627 (cit. on p. 713).

Michael Carl, Max Pumperla, and Bernd Siebert (2010). "A tropical view of Landau–Ginzburg models" (cit. on p. 711).

Vyjayanthi Chari and Andrew Pressley (1995). *A guide to quantum groups*. Corrected reprint of the 1994 original. Cambridge University Press, Cambridge, pp. xvi+651. MR: 1358358 (cit. on p. 713).

Man Wai Cheung, Mark Gross, Greg Muller, Gregg Musiker, Dylan Rupel, Salvatore Stella, and Harold Williams (2017). "The greedy basis equals the theta basis: a rank two haiku". *J. Combin. Theory Ser. A* 145, pp. 150–171. MR: 3551649 (cit. on p. 712).

Cheol-Hyun Cho and Yong-Geun Oh (2006). "Floer cohomology and disc instantons of Lagrangian torus fibers in Fano toric manifolds". *Asian J. Math.* 10.4, pp. 773–814. MR: 2282365 (cit. on pp. 699, 700, 707).

Pierre Deligne (1971). "Théorie de Hodge. II". *Inst. Hautes Études Sci. Publ. Math.* 40, pp. 5–57. MR: 0498551 (cit. on pp. 691, 693).

Tommaso de Fernex, János Kollár, and Chenyang Xu (Dec. 2012). "The dual complex of singularities". arXiv: 1212.1675 (cit. on pp. 691, 695).

Vladimir Fock and Alexander Goncharov (2006). "Moduli spaces of local systems and higher Teichmüller theory". *Publ. Math. Inst. Hautes Études Sci.* 103, pp. 1–211. MR: 2233852 (cit. on pp. 690, 702).

Sergey Fomin and Andrei Zelevinsky (2002). "Cluster algebras. I. Foundations". *J. Amer. Math. Soc.* 15.2, pp. 497–529. MR: 1887642 (cit. on p. 690).

– (2003). "Cluster algebras. II. Finite type classification". *Invent. Math.* 154.1, pp. 63–121. MR: 2004457 (cit. on p. 712).

William Fulton (1993). *Introduction to toric varieties*. Vol. 131. Annals of Mathematics Studies. The William H. Roever Lectures in Geometry. Princeton University Press, Princeton, NJ, pp. xii+157. MR: 1234037 (cit. on pp. 693, 695).

William Fulton and Joe Harris (1991). *Representation theory*. Vol. 129. Graduate Texts in Mathematics. A first course, Readings in Mathematics. Springer-Verlag, New York, pp. xvi+551. MR: 1153249 (cit. on p. 713).

Sheel Ganatra (2012). *Symplectic Cohomology and Duality for the Wrapped Fukaya Category*. Thesis (Ph.D.)–Massachusetts Institute of Technology. ProQuest LLC, Ann Arbor, MI, (no paging). arXiv: 1304.7312. MR: 3121862 (cit. on p. 702).

Michael Gekhtman, Michael Shapiro, and Alek Vainshtein (2010). *Cluster algebras and Poisson geometry*. Vol. 167. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. xvi+246. MR: 2683456 (cit. on p. 713).

Alexander Goncharov and Linhui Shen (2015). "Geometry of canonical bases and mirror symmetry". *Invent. Math.* 202.2, pp. 487–633. MR: 3418241 (cit. on pp. 713, 714).

Mark Gross, Paul Hacking, and Sean Keel (2015a). "Birational geometry of cluster algebras". *Algebr. Geom.* 2.2, pp. 137–175. MR: 3350154 (cit. on pp. 690, 691, 694, 708).

– (2015b). "Mirror symmetry for log Calabi-Yau surfaces I". *Publ. Math. Inst. Hautes Études Sci.* 122, pp. 65–168. MR: 3415066 (cit. on pp. 692, 704).

Mark Gross, Paul Hacking, Sean Keel, and Maxim Kontsevich (2018). "Canonical bases for cluster algebras". *J. Amer. Math. Soc.* 31.2, pp. 497–608. arXiv: 1411.1394. MR: 3758151 (cit. on pp. 706, 709–714).

Mark Gross, Rahul Pandharipande, and Bernd Siebert (2010). "The tropical vertex". *Duke Math. J.* 153.2, pp. 297–362. MR: 2667135 (cit. on pp. 706, 707).

Mark Gross and Bernd Siebert (2011). "From real affine geometry to complex geometry". *Ann. of Math. (2)* 174.3, pp. 1301–1428. MR: 2846484 (cit. on p. 706).

– (Sept. 2016). "Intrinsic mirror symmetry and punctured Gromov-Witten invariants". arXiv: 1609.00624 (cit. on p. 707).

V. A. Iskovskih and Ju. I. Manin (1971). "Three-dimensional quartics and counterexamples to the Lüroth problem". *Mat. Sb. (N.S.)* 86(128), pp. 140–166. MR: 0291172 (cit. on p. 692).

Ailsa Keating (Oct. 2015). "Homological mirror symmetry for hypersurface cusp singularities". arXiv: 1510.08911 (cit. on p. 704).

János Kollár and Shigefumi Mori (1998). *Birational geometry of algebraic varieties*. Vol. 134. Cambridge Tracts in Mathematics. With the collaboration of C. H. Clemens and A. Corti, Translated from the 1998 Japanese original. Cambridge University Press, Cambridge, pp. viii+254. MR: 1658959 (cit. on pp. 689, 696).

János Kollár and Chenyang Xu (2016). "The dual complex of Calabi-Yau pairs". *Invent. Math.* 205.3, pp. 527–557. MR: 3539921 (cit. on p. 692).

Maxim Kontsevich (1995). "Homological algebra of mirror symmetry". In: *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*. Birkhäuser, Basel, pp. 120–139. MR: 1403918 (cit. on p. 698).

Maxim Kontsevich and Yan Soibelman (2001). "Homological mirror symmetry and torus fibrations". In: *Symplectic geometry and mirror symmetry (Seoul, 2000)*. World Sci. Publ., River Edge, NJ, pp. 203–263. MR: 1882331 (cit. on p. 698).

– (2006). "Affine structures and non-Archimedean analytic spaces". In: *The unity of mathematics*. Vol. 244. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 321–385. MR: 2181810 (cit. on pp. 695, 706).

Timothy Magee (Sept. 2017). "GHK mirror symmetry, the Knutson-Tao hive cone, and Littlewood-Richardson coefficients". arXiv: 1709.05776 (cit. on p. 713).

D. Mumford, J. Fogarty, and F. Kirwan (1994). *Geometric invariant theory*. 3rd ed. Vol. 34. Ergebnisse der Mathematik und ihrer Grenzgebiete (2). Springer-Verlag, Berlin, pp. xiv+292. MR: 1304906 (cit. on p. 697).

James Pascaleff (Apr. 2013). "On the symplectic cohomology of log Calabi-Yau surfaces". arXiv: 1304.5298 (cit. on p. 702).

Alexander Polishchuk (2003). *Abelian varieties, theta functions and the Fourier transform*. Vol. 153. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, pp. xvi+292. MR: 1987784 (cit. on p. 698).

Paul Seidel (2002). “Fukaya categories and deformations”. In: *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*. Higher Ed. Press, Beijing, pp. 351–360. MR: 1957046 (cit. on p. 702).

– (2008). “A biased view of symplectic cohomology”. In: *Current developments in mathematics, 2006*. Int. Press, Somerville, MA, pp. 211–253. MR: 2459307 (cit. on p. 702).

– (2013). “Lectures on categorical dynamics and symplectic topology” (cit. on p. 701).

Andrew Strominger, Shing-Tung Yau, and Eric Zaslow (1996). “Mirror symmetry is T-duality”. *Nuclear Phys. B* 479.1-2, pp. 243–259. MR: 1429831 (cit. on pp. 690, 696).

Received 2017-12-06.

PAUL HACKING
DEPARTMENT OF MATHEMATICS AND STATISTICS
LEDERLE GRADUATE RESEARCH TOWER
UNIVERSITY OF MASSACHUSETTS
AMHERST, MA 01003-9305
hacking@math.umass.edu

SEAN KEEL
DEPARTMENT OF MATHEMATICS
1 UNIVERSITY STATION C1200
AUSTIN TX 78712-0257
keel@math.utexas.edu

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (717–736)

ALGEBRAIC SURFACES WITH MINIMAL BETTI NUMBERS

JONGHAE KEUM (금종해)

Abstract

These are algebraic surfaces with the Betti numbers of the complex projective plane, and are called $\mathbb{Q}$-homology projective planes. Fake projective planes and the complex projective plane are smooth examples. We describe recent progress in the study of such surfaces, singular ones and fake projective planes. We also discuss open questions.

1 $\mathbb{Q}$-homology Projective Planes and Montgomery-Yang problem

A normal projective surface with the Betti numbers of the complex projective plane $\mathbb{CP}^2$ is called a *rational homology projective plane* or a *$\mathbb{Q}$-homology* $\mathbb{CP}^2$. When a normal projective surface S has only rational singularities, S is a $\mathbb{Q}$-homology $\mathbb{CP}^2$ if its second Betti number $b_2(S) = 1$. This can be seen easily by considering the Albanese fibration on a resolution of S.

It is known that a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with quotient singularities (and no worse singularities) has at most 5 singular points (cf. Hwang and Keum [2011b, Corollary 3.4]). The $\mathbb{Q}$-homology projective planes with 5 quotient singularities were classified in Hwang and Keum [ibid.].

In this section we summarize progress on the Algebraic Montgomery-Yang problem, which was formulated by J. Kollár.

Conjecture 1.1 (Algebraic Montgomery–Yang Problem Kollár [2008]). *Let S be a $\mathbb{Q}$-homology projective plane with quotient singularities. Assume that $S^0 := S \backslash Sing(S)$ is simply connected. Then S has at most* 3 *singular points.*

This is the algebraic version of Montgomery–Yang Problem Fintushel and Stern [1987], which was originated from pseudofree circle group actions on higher dimensional sphere.

MSC2010: primary 14J29; secondary 14F05, 14J17, 14J26, 32Q40, 32N15.
Keywords: fake projective planes, ball quotients, $\mathbb{Q}$-homology projective planes, singular 4-manifolds, Montgomery–Yang problem.

Pseudofree circle group actions are those that have no points fixed by the entire circle group but that have isolated circles that are pointwise fixed by finite cyclic subgroups.

- Work over the field $\mathbb{C}$ of complex numbers, except a few remarks in positive characteristic and in differentiable case.

Definition 1.2. A normal projective surface S is called a $\mathbb{Q}$-homology $\mathbb{CP}^2$ if it has the same Betti numbers as $\mathbb{CP}^2$, i.e. $(b_0, b_1, b_2, b_3, b_4) = (1, 0, 1, 0, 1)$.

Examples 1.3.

1. If S is smooth, then $S = \mathbb{CP}^2$ or a fake projective plane(fpp).
2. If S is singular and has only A_1-singularities, then S is isomorphic to the quadric cone $(xy = z^2)$ in $\mathbb{CP}^3$, i.e., the weighted projective plane $\mathbb{CP}^2(1, 1, 2)$ (cf. Keum [2010]).
3. Cubic surfaces with $3A_2$ in $\mathbb{CP}^3$ (such surfaces can be shown to be isomorphic to $(w^3 = xyz)$).
4. If S has A_2-singularities only, then S has $3A_2$ or $4A_2$ and $S = \mathbb{CP}^2/G$ or fpp$/G$, where $G \cong \mathbb{Z}/3$ or $(\mathbb{Z}/3)^2$.
5. If S has A_1 or A_2-singularities only, then $S = \mathbb{CP}^2(1, 2, 3)$ or one of the above (see Keum [2015] for details).

In this section, S has at worst quotient singularities. Then it is easy to see that

- S is a $\mathbb{Q}$-homology $\mathbb{CP}^2$ if and only if $b_2(S) = 1$.
- A minimal resolution of S has $p_g = q = 0$.

1.1 Trichotomy: K_S ample, $-$ample, numerically trivial. Let S a $\mathbb{Q}$-hom $\mathbb{CP}^2$ with quotient singularities, $f : S' \to S$ a minimal resolution. The canonical class K_S falls in one of the three cases.

1. $-K_S$ is ample (log del Pezzo surfaces of Picard number 1): their minimal resolutions have Kodaira dimension $\kappa(S') = -\infty$; typical examples are $\mathbb{CP}^2/G$, $\mathbb{CP}^2(a, b, c)$.
2. K_S is numerically trivial (log Enriques surfaces of Picard number 1): their minimal resolutions have Kodaira dimension $\kappa(S') = -\infty, 0$.

3. K_S is ample: : their minimal resolutions have Kodaira dimension $\kappa(S') = -\infty, 0, 1, 2$; typical examples are fake projective planes and their quotients, suitable contraction of a suitable blowup of some Enriques surface or $\mathbb{CP}^2$.

The following problem was raised by J. Kollár [2008].

Problem 1.4. Classify all $\mathbb{Q}$-homology projective planes with quotient singularities.

1.2 The Maximum Number of Quotient Singularities. Let S be a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with quotient singularities. From the orbifold Bogomolov–Miyaoka–Yau inequality (Sakai [1980], Miyaoka [1984], and Megyesi [1999]), one can derive that S has at most 5 singular points (see Hwang and Keum [2011b, Corollary 3.4]). Many examples with 4 or less singular points were provided (Brenton [1977] and Brenton, Drucker, and Prins [1981]), but no example with 5 singular points until D. Hwang and the author characterised the case with 5 singular points.

Theorem 1.5 (Hwang and Keum [2011b]). *Let S be a $\mathbb{Q}$-homology projective plane with quotient singularities. Then S has at most 4 singular points except the following case which is supported by an example: S has 5 singular points of type $3A_1 + 2A_3$.*
In the exceptional case the minimal resolution of S is an Enriques surface.

In fact, our proof assumed the condition that K_S is nef to apply the orbifold Bogomolov–Miyaoka–Yau inequality. On the other hand, the case where $-K_S$ is ample was dealt with by Belousov [2008, 2009]. He proved that log del Pezzo surfaces of Picard number 1 with quotient singularities have at most 4 singular points.

Corollary 1.6. *The following hold true.*

1. *$\mathbb{Q}$-cohomology projective planes with quotient singularities have at most 4 singular points except the case given in the above theorem.*

2. *$\mathbb{Z}$-homology projective planes with quotient singularities have at most 4 singular points.*

Here, a $\mathbb{Q}$-cohomology projective plane is a normal projective complex surface having the same $\mathbb{Q}$-cohomology ring with $\mathbb{CP}^2$. A $\mathbb{Q}$-homology projective plane with quotient singularities is a $\mathbb{Q}$-cohomology projective plane.

The problem of determining the maximum number of singular points on $\mathbb{Q}$-homology projective planes with quotient singularities is related to the algebraic Montgomery–Yang problem (Montgomery and Yang [1972] and Kollár [2008]).

Remark 1.7. (1) Every $\mathbb{Z}$-cohomology $\mathbb{CP}^2$ with quotient singularities has at most 1 singular point, and, if it has, then the singularity must be of type E_8 (Bindschadler and Brenton [1984]).

(2) If a $\mathbb{Q}$-homology projective plane S is allowed to have rational singular points (quotient singular points are rational, but the converse does not holds), then there is no bound for the number of singular points. In fact, there are $\mathbb{Q}$-homology projective planes with an arbitrary number of rational singular points. Such examples can be constructed by modifying Example 5 from Kollár [2008]: take a minimal ruled surface $\mathbb{F}_e \to \mathbb{P}^1$ with negative section E, blow up m distinct fibres into m strings of 3 rational curves (-2)—(-1)—(-2), then contract the proper transform of E with the m adjacent (-2)-curves, and also the m remaining (-2)-curves. If $e = -E^2$ is sufficiently larger than m, then the big singular point is rational and yields a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with $m+1$ rational singular points.

(3) In char 2 there is a rational example of Picard number 1 with seven A_1-singularities.

Theorem 1.8 (The orbifold BMY inequality)**.** *Let S be a normal projective surface with quotient singularities. Assume that K_S is nef. Then*

$$K_S^2 \leq 3e_{orb}(S).$$

Theorem 1.9 (The weak oBMY inequality, Keel and McKernan [1999])**.** *Let S be a normal projective surface with quotient singularities. Assume that $-K_S$ is nef. Then*

$$0 \leq e_{orb}(S).$$

Here the orbifold Euler characteristic is defined by

$$e_{orb}(S) := e(S) - \sum_{p \in Sing(S)} \left(1 - \frac{1}{|\pi_1(L_p)|}\right)$$

1.3 Smooth $\mathbf{S}^1$-actions on $\mathbf{S}^m$. Let $\mathbf{S}^1$ be the circle group. Consider a faithful $\mathcal{C}^\infty$ action of $\mathbf{S}^1$ on the m-dimensional sphere $\mathbf{S}^m$

$$\mathbf{S}^1 \subset Diff(\mathbf{S}^m).$$

The identity element $1 \in \mathbf{S}^1$ acts as the identity on $\mathbf{S}^m$. Each diffeomorphism $g \in \mathbf{S}^1$ is homotopic to the identity on $\mathbf{S}^m$. By Lefschetz Fixed Point Formula,

$$e(Fix(g)) = e(Fix(1)) = e(\mathbf{S}^m).$$

If m is even, then $e(\mathbf{S}^m) = 2$ and such an action has fixed points, so is not pseudofree. Pseudofree circle actions are those that have no points fixed by the entire circle group but that have isolated circles that are pointwise fixed by finite cyclic subgroups.

Assume $m = 2n - 1$ odd.

Definition 1.10. A faithful $\mathcal{C}^\infty$-action of $\mathbf{S}^1$ on $\mathbf{S}^{2n-1}$

$$\mathbf{S}^1 \times \mathbf{S}^{2n-1} \to \mathbf{S}^{2n-1}$$

is called *pseudofree* if it is free except for finitely many orbits whose isotropy groups $\mathbb{Z}/a_1\mathbb{Z}, \ldots, \mathbb{Z}/a_k\mathbb{Z}$ have pairwise prime orders.

1.4 Pseudofree $\mathbf{S}^1$-actions on $\mathbf{S}^{2n-1}$.

Example 1.11 (Linear action)**.** For $a_1, ..., a_n$ pairwise prime

$$\mathbf{S}^1 \times \mathbf{S}^{2n-1} \to \mathbf{S}^{2n-1}$$

$$(\lambda, (z_1, z_2, ..., z_n)) \mapsto (\lambda^{a_1} z_1, \lambda^{a_2} z_2, ..., \lambda^{a_n} z_n)$$

where

$$\mathbf{S}^{2n-1} = \{(z_1, z_2, ..., z_n) : |z_1|^2 + |z_2|^2 + ... + |z_n|^2 = 1\} \subset \mathbb{C}^n$$
$$\mathbf{S}^1 = \{\lambda : |\lambda| = 1\} \subset \mathbb{C}.$$

In such a linear action

- $$\mathbf{S}^{2n-1}/\mathbf{S}^1 \cong \mathbb{CP}^{n-1}(a_1, a_2, ..., a_n);$$
- The orbit of the i-th coordinate point $e_i = (0, ..., 0, 1, 0..., 0) \in \mathbf{S}^{2n-1}$ is an exceptional orbit iff $a_i \geq 2$;
- The orbit of a non-coordinate point of $\mathbf{S}^{2n-1}$ is NOT exceptional;
- This action has at most n exceptional orbits;
- The quotient map $\mathbf{S}^{2n-1} \to \mathbb{CP}^{n-1}(a_1, a_2, ..., a_n)$ is a Seifert fibration.

For Seifert fibrations we recall the following.

1. For $n = 2$ Seifert [1933] showed that every pseudofree $\mathbf{S}^1$-action on $\mathbf{S}^3$ is diffeomorphic to a diagonal one and hence has at most 2 exceptional orbits.

2. For $n = 4$ Montgomery and Yang [1972] showed that given arbitrary collection of pairwise prime positive integers $a_1, \ldots, a_k$, there is a pseudofree $\mathbf{S}^1$-action on a homotopy $\mathbf{S}^7$ whose exceptional orbits have exactly those orders.

3. Petrie [1975] generalised the result of Montgomery-Yang for all $n \geq 5$.

Conjecture 1.12 (Montgomery–Yang problem, Fintushel and Stern [1987])**.** *A pseudo-free $\mathbf{S}^1$-action on $\mathbf{S}^5$ has at most 3 exceptional orbits.*

The problem has remained unsolved since its formulation.

- Pseudo-free $\mathbf{S}^1$-actions on a manifold Σ have been studied in terms of the pseudofree orbifold $\Sigma/\mathbf{S}^1$ (see e.g., Fintushel and Stern [1985, 1987]).
- The orbifold $X = \mathbf{S}^5/\mathbf{S}^1$ is a 4-manifold with isolated singularities whose neighborhoods are cones over lens spaces corresponding to the exceptional orbits of the $\mathbf{S}^1$-action.
- Easy to check that X is simply connected and $H_2(X, \mathbb{Z})$ has rank 1 and intersection matrix $(1/a_1 a_2 \cdots a_k)$.
- An exceptional orbit with isotropy type $\mathbb{Z}/a$ has an equivariant tubular neighborhood which may be identified with $\mathbb{C} \times \mathbb{C} \times \mathbf{S}^1$ with a $\mathbf{S}^1$-action

$$\lambda \cdot (z, w, u) = (\lambda^r z, \lambda^s w, \lambda^a u)$$

  where r and s are relatively prime to a.

The following 1-1 correspondence was known to Montgomery–Yang, Fintushel–Stern, and revisited by Kollár [2005, 2008].

Theorem 1.13. *There is a one-to-one correspondence between:*

1. *Pseudofree $\mathbf{S}^1$-actions on $\mathbb{Q}$-homology 5-spheres Σ with $H_1(\Sigma, \mathbb{Z}) = 0$.*
2. *Compact differentiable 4-manifolds M with boundary such that*
   - *(a)* $\partial M = \bigcup_i L_i$ *is a disjoint union of lens spaces* $L_i = \mathbf{S}^3/\mathbb{Z}_{a_i}$,
   - *(b) the orders a_i's are pairwise prime,*
   - *(c)* $H_1(M, \mathbb{Z}) = 0$,
   - *(d)* $H_2(M, \mathbb{Z}) \cong \mathbb{Z}$.

Furthermore, Σ is diffeomorphic to $\mathbf{S}^5$ iff $\pi_1(M) = 1$.

1.5 Algebraic Montgomery-Yang Problem. This is the Montgomery-Yang Problem when the pseudofree orbifold $\mathbf{S}^5/\mathbf{S}^1$ attains a structure of a normal projective surface.

Conjecture 1.14 (Kollár [2008]). *Let S be a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with at worst quotient singularities. If the smooth part S^0 has $\pi_1(S^0) = \{1\}$, then S has at most 3 singular points.*

What happens if the condition $\pi_1(S^0) = \{1\}$ is replaced by the weaker condition $H_1(S^0, \mathbb{Z}) = 0$?

Remark 1.15. If $H_1(S^0, \mathbb{Z}) = 0$, then
(1) K_S cannot be numerically trivial;
(2) it follows from the oBMY that $|Sing(S)| \leq 4$.

There are infinitely many examples S with

$$H_1(S^0, \mathbb{Z}) = 0, \quad \pi_1(S^0) \neq \{1\}, \quad |Sing(S)| = 4,$$

obtained from the classification of surface quotient singularities by Brieskorn [1967/1968].

Example 1.16 (Brieskorn quotients). Let $I_m \subset GL(2, \mathbb{C})$ be the $2m$-ary icosahedral group $I_m = \mathbb{Z}_{2m}.\mathfrak{A}_5$,

$$1 \to \mathbb{Z}_{2m} \to I_m \to \mathfrak{A}_5 \subset PSL(2, \mathbb{C}).$$

The action of I_m on $\mathbb{C}^2$ extends naturally to $\mathbb{CP}^2$. Then

$$S := \mathbb{CP}^2/I_m$$

is a $\mathbb{Q}$-homology $\mathbb{CP}^2$ such that

- $-K_S$ ample;
- S has 4 quotient singularities: one non-cyclic singularity of type I_m (the image of the origin $O \in \mathbb{C}^2$) and 3 cyclic singularities of order $2, 3, 5$ (on the image of the line at infinity);
- $\pi_1(S^0) = \mathfrak{A}_5$, hence $H_1(S^0, \mathbb{Z}) = 0$.

1.6 Progress on Algebraic Montgomery-Yang Problem. Algebraic Montgomery-Yang problem holds true if S has at least 1 non-cyclic singular point.

Theorem 1.17 (Hwang and Keum [2011a]). *Let S be a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with quotient singular points, not all cyclic, such that $\pi_1(S^0) = \{1\}$. Then $|Sing(S)| \leq 3$.*

More precisely

Theorem 1.18. *Let S be a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with 4 or more quotient singular points, not all cyclic, such that $H_1(S^0, \mathbb{Z}) = 0$. Then S is isomorphic to a Brieskorn quotient.*

Remark 1.19. In Hwang and Keum [2011a], though the proof was correct, a wrong conclusion was made in the statement (4) of Theorem 3 that the smooth part S^0 of such a surface S is deformation equivalent to the smooth part of a Brieskorn quotient. A corrected statement was given in Hwang and Keum [2009].

More Progress on the Algebraic Montgomery-Yang Problem:

Theorem 1.20 (Hwang and Keum [2013, 2014])**.** *Let S be a $\mathbb{Q}$-homology $\mathbb{CP}^2$ with cyclic singularities such that $H_1(S^0, \mathbb{Z}) = 0$. If either S is not rational or $-K_S$ is ample, then $|Sing(S)| \leq 3$.*

The Remaining Case of the Algebraic Montgomery-Yang Problem:

S is a $\mathbb{Q}$-homology $\mathbb{CP}^2$ such that

1. S has cyclic singularities only,
2. S is a rational surface with K_S ample.

Looking at the adjunction formula

$$K_{S'} = \pi^* K_S - \sum D_p,$$

where $S' \to S$ is a resolution, one sees that K_S, though ample, is "smaller than $\sum D_p$" so that no positive multiple of $K_{S'}$ is effective. Such surfaces were given by

Keel and McKernan [1999];
Kollár [2008]: infinite series of examples with $|Sing(S)| = 2$;
Hwang and Keum [2012]: infinite series of examples with $|Sing(S)| = 1, 2, 3$.
Urzúa and Yáñez [2016]: characterization of Kollár surfaces.

There is no known example with $|Sing(S)| = 4$, even if the condition $H_1(S^0, \mathbb{Z}) = 0$ is removed.

Problem 1.21. Is there a $\mathbb{Q}$-homology projective plane S which is a rational surface with K_S ample and $|Sing(S)| = 4$?

Another interesting line of research is to obtain surfaces with quotient singularities with small volume. Alexeev and Liu [2016] has constructed a surface S with log terminal singularities (quotient singularities) and ample canonical class that has $K_S^2 = 1/48983$ and a log canonical pair (S, B) with a nonempty reduced divisor B and ample $K_S + B$ that has $(K_S + B)^2 = 1/462$, both examples signifcantly improve known record.

1.7 Cascade structure on rational $\mathbb{Q}$-homology projective planes. A rational $\mathbb{Q}$-homology $\mathbb{CP}^2$ is obtained by blow-ups and downs from a "basic surface" which is a rational minimal surface with certain configuration of curves.
In the case where $-K_S$ ample, all basic surfaces have been classified by Hwang [n.d.].

1.8 Gorenstein $\mathbb{Q}$-homology projective planes. These are $\mathbb{Q}$-homology projective planes with ADE-singular points (i.e., rational double points).

Let R be the singularity type, i.e., the corresponding root sublattice of the cohomology lattice of S', the minimal resolution of S.
Since S is Gorenstein, rank(R) is bounded.

$$1 + \text{rank}(R) = b_2(S') = 10 - K_{S'}^2 = 10 - K_S^2 \leq 10,$$

$$\text{rank}(R) \leq 9$$

with equality iff K_S is numerically trivial iff S' is an Enriques surface.

With D. Hwang and H. Ohashi, we classified all possible singularity types of Gorenstein $\mathbb{Q}$-homology projective planes. There are 58 types total.

Theorem 1.22 (Hwang, Keum, and Ohashi [2015]). *The singularity type R of a Gorenstein $\mathbb{Q}$-homology $\mathbb{CP}^2$ is one of the following:*

(1) K_S *±ample* (27 *types*)*:*
A_8, A_7, D_8, E_8, E_7, E_6, D_5, A_4, A_1;
$A_7 \oplus A_1$, $A_5 \oplus A_2$, $A_5 \oplus A_1$, $2A_4$, $A_2 \oplus A_1$, $D_6 \oplus A_1$, $D_5 \oplus A_3$, $2D_4$, $E_7 \oplus A_1$, $E_6 \oplus A_2$;
$A_5 \oplus A_2 \oplus A_1$, $2A_3 \oplus A_1$, $A_3 \oplus 2A_1$, $3A_2$, $D_6 \oplus 2A_1$;
$2A_3 \oplus 2A_1$, $4A_2$, $D_4 \oplus 3A_1$,

(2) K_S *numerically trivial* (31 *types*)
A_9, D_9;
$A_8 \oplus A_1$, $A_7 \oplus A_2$, $A_5 \oplus A_4$, $D_8 \oplus A_1$, $D_6 \oplus A_3$, $D_5 \oplus A_4$, $D_5 \oplus D_4$, $E_8 \oplus A_1$, $E_7 \oplus A_2$, $E_6 \oplus A_3$;
$A_7 \oplus 2A_1$, $A_6 \oplus A_2 \oplus A_1$, $A_5 \oplus A_3 \oplus A_1$, $A_5 \oplus 2A_2$, $2A_4 \oplus A_1$, $3A_3$, $D_7 \oplus 2A_1$, $D_6 \oplus A_2 \oplus A_1$, $D_5 \oplus A_3 \oplus A_1$, $2D_4 \oplus A_1$, $E_7 \oplus 2A_1$, $E_6 \oplus A_2 \oplus A_1$;

$A_5 \oplus A_2 \oplus 2A_1$, $A_4 \oplus A_3 \oplus 2A_1$, $2A_3 \oplus A_2 \oplus A_1$, $A_3 \oplus 3A_2$, $D_6 \oplus 3A_1$, $D_4 \oplus A_3 \oplus 2A_1$; $2A_3 \oplus 3A_1$.

The 27 types with $-K_S$ ample were classified by Furushima [1986], Miyanishi and Zhang [1988], Ye [2002]. Our method uses only lattice theory, different from theirs.

Among the 31 types with $K_S \equiv 0$, 29 types are supported by Enriques surfaces with finite automorphism group. Enriques surfaces with $|Aut| < \infty$ must have finitely many (-2)-curves, and were classified by Nikulin and Kondō [1986] into 7 families, two families 1-dimensional and five 0-dimensional. Schütt [2015] has constucted explicitly, for each of the 31 types, the moduli space of such Enriques surfaces, all 1-dimensional.

Remark 1.23. In positive characteristic the case with $K_S \equiv 0$ has been classified by M. Schütt [2016, 2017], which build on recent deep results of Katsura and Kondō [2018], Martin [2017], and Katsura, Kondō, and Martin [2017], also on a unpublished work of Dolgachev-Liedtke.

1.9 The differentiable case. Let M be a smooth, compact 4-manifold whose boundary components are spherical, that is, lens spaces $L_i = \mathbf{S}^3$. One can then attach cones to each boundary component to get a 4-dimensional orbifold S. As in the algebraic case, there is a minimal resolution $f : S' \to S$, where S' is a smooth, compact 4-manifold without boundary.

To each singular point $p \in S$ (the vertex of each cone), we assign a uniquely defined class $\mathfrak{D}_p = \sum(a_j E_j) \in H^2(S', \mathbb{Q})$ such that

$$\mathfrak{D}_p \cdot E_i = 2 + E_i^2$$

for each component E_i of $f^{-1}(p)$. We always assume that S and S' satisfy the following two conditions:

(1) S is a $\mathbb{Q}$-homology $\mathbb{CP}^2$, i.e. $H^1(S, \mathbb{Q}) = 0$ and $H^2(S, \mathbb{Q}) \cong \mathbb{Q}$.

(2) The intersection form on $H^2(S', \mathbb{Q})$ is indefinite, and is negative definite on the subspace generated by the classes of the exceptional curves of f.

If there is a class $K_{S'} \in H^2(S', \mathbb{Q})$ satisfying both the Noether formula

$$K_{S'}^2 = 10 - b_2(S')$$

and the adjunction formula

$$K_{S'} \cdot E + E^2 = -2$$

for each exceptional curve E of f, we call it a *formal canonical class of* S', and the class $K_{S'} + \sum \mathfrak{D}_p \in H^2(S', \mathbb{Q})$ a *formal canonical class of* S.

Theorem 1.24 (Hwang and Keum [2011b] Theorem 8.1)**.** *Let M, S, and S' be the same as above satisfying the conditions* (1) *and* (2)*. Assume that S' admits a formal canonical class $K_{S'}$. Assume further that*

$$K_{S'}^2 - \sum_{p \in Sing(S)} \mathfrak{D}_p^2 \leq 3e_{orb}(S).$$

Then M has at most 4 boundary components except the following two cases: M has 5 boundary components of type $3A_1 + 2A_3$ or $4A_1 + D_5$.

Note that the assumptions in Theorem 1.24 all hold for algebraic $\mathbb{Q}$-homology projective planes with quotient singularities such that the canonical divisor is nef.

Theorem 1.25 (Hwang and Keum [ibid.] Theorem 8.2)**.** *Let M, S, and S' be the same as above satisfying the conditions* (1) *and* (2)*. Assume that S' admits a formal canonical class $K_{S'}$. Assume further that*

$$0 \leq e_{orb}(S).$$

Then M has at most 5 boundary components. The bound is sharp.

The assumptions in Theorem 1.25 all hold for algebraic $\mathbb{Q}$-homology projective planes with quotient singularities.

1.10 The symplectic case. If S is a symplectic orbifold, then S' is a symplectic manifold and the symplectic canonical class $K_{S'}$ gives a formal canonical class.

Problem 1.26. Is there a Bogomolov–Miyaoka–Yau type inequality for symplectic 4-manifolds?
Is there an orbifold Bogomolov–Miyaoka–Yau type inequality for symplectic orbifolds?

The following question is also interesting in view of Sasakian geometry.

Problem 1.27 (Muñoz, Rojo, and Tralle [2016])**.** There does not exist a Kähler manifold or a Kähler orbifold with $b_1 = 0$ and $b_2 \geq 2$ having b_2 disjoint complex curves all of genus $g \geq 1$ which generate $H_2(S, \mathbb{Q})$.

A ruled surface over a curve of genus g has two disjoint curves, the negative section and a general section, of genus g, but has $b_1 \neq 0$ if $g \geq 1$.

2 Fake projective planes

A compact complex surface with the same Betti numbers as the complex projective plane is called *a fake projective plane* if it is not biholomorphic to the complex projective plane.

A fake projective plane has ample canonical divisor, so it is a smooth proper (geometrically connected) surface of general type with geometric genus $p_g = 0$ and self-intersection of canonical class $K^2 = 9$ (this definition extends to arbitrary characteristic.) The existence of a fake projective plane was first proved by Mumford [1979] based on the theory of 2-adic uniformization, and later two more examples by Ishida and Kato [1998] 1998) in a similar method. Keum [2006] gave a construction of a fake projective plane with an order 7 automorphism, which is birational to an order 7 cyclic cover of a Dolgachev surface. This surface and Mumford fake projective plane belong to the same class, in the sense that their fundamental groups are both contained in the same maximal arithmetic subgroup of the isometry group of the complex 2-ball.

Fake projective planes have Chern numbers $c_1^2 = 3c_2 = 9$ and are complex 2-ball quotients by Aubin [1976] and Yau [1977]. Such ball quotients are strongly rigid by Mostow's rigidity theorem (Mostow [1973]), that is, determined by fundamental group up to holomorphic or anti-holomorphic isomorphism. Fake projective planes come in complex conjugate pairs by Kulikov and Kharlamov [2002] and have been classified as quotients of the two-dimensional complex ball by explicitly written co-compact torsion-free arithmetic subgroups of $\mathrm{PU}(2,1)$ by Prasad and Yeung [2007, 2010] and Cartwright and Steger [2010, n.d.]. The arithmeticity of their fundamental groups was proved by Klingler [2003]. There are exactly 100 fake projective planes total, corresponding to 50 distinct fundamental groups. Cartwright and Steger also computed the automorphism group of each fake projective plane X, which is given by $\mathrm{Aut}(X) \cong N(X)/\pi_1(X)$, where $N(X)$ is the normalizer of $\pi_1(X)$ in its maximal arithmetic subgroup of $\mathrm{PU}(2,1)$. In particular $\mathrm{Aut}(X) \cong \{1\}$, $\mathbb{Z}_3$, $\mathbb{Z}_3^2$ or G_{21} where $\mathbb{Z}_n$ is the cyclic group of order n and G_{21} is the unique non-abelian group of order 21. Among the 50 pairs exactly 33 admit non-trivial automorphisms: 3 pairs have $\mathrm{Aut} \cong G_{21}$, 3 pairs have $\mathrm{Aut} \cong \mathbb{Z}_3^2$ and 27 pairs have $\mathrm{Aut} \cong \mathbb{Z}_3$.

For each pair of fake projective planes Cartwright and Steger [n.d.] also computed the torsion group

$$H_1(X,\mathbb{Z}) = \mathrm{Tor}(H^2(X,\mathbb{Z})) = \mathrm{Tor}(\mathrm{Pic}(X))$$

which is the abelianization of the fundamental group. According to their computation exactly 29 pairs of fake projective planes have a 3-torsion in $H_1(X,\mathbb{Z})$.

In this section we summarize recent progress on these fascinating objects.

2.1 Picard group of a fake projective plane. Since $p_g(X) = q(X) = 0$, the long exact sequence induced by the exponential sequence gives $\mathrm{Pic}(X) = H^2(X,\mathbb{Z})$. By the universal coefficient theorem, $\mathrm{Tor}H^2(X,\mathbb{Z}) = \mathrm{Tor}H_1(X,\mathbb{Z})$. This implies that

$$\mathrm{Pic}(X) = H^2(X,\mathbb{Z}) \cong \mathbb{Z} \times H_1(X,\mathbb{Z}).$$

Two ample line bundles with the same self-intersection number on a fake projective plane differ by a torsion.

It can be shown (cf. Keum [2017, Lemma 1.5]) that if a fake projective plane X has no 3-torsion in $H_1(X,\mathbb{Z})$ (21 pairs of fake projective planes satisfy this property), then the canonical class K_X is divisible by 3 and has a unique cube root, i.e., a unique line bundle L_0 up to isomorphism such that $3L_0 \equiv K_X$.

By a result of Kollár [1995, p. 96] the 3-divisibility of K_X is equivalent to the liftability of the fundamental group to $\mathrm{SU}(2,1)$. Except 4 pairs of fake projective planes the fundamental groups lift to $\mathrm{SU}(2,1)$ (Prasad and Yeung [2010, Section 10.4], Cartwright and Steger [2010, n.d.]). In the notation of Cartwright and Steger [2010], these exceptional 4 pairs are the 3 pairs in the class $(\mathcal{C}_{18}, p=3, \{2\})$, whose automorphism groups are of order 3, and the one in the class $(\mathcal{C}_{18}, p=3, \{2I\})$, whose automorphism group is trivial. There are fake projective planes with a 3-torsion and with canonical class divisible by 3 Cartwright and Steger [n.d.].

2.2 Quotients of fake projective planes. Let X be a fake projective plane with a nontrivial group G acting on it. In Keum [2008], all possible structures of the quotient surface X/G and its minimal resolution were classified:

Theorem 2.1 (Keum [ibid.]). *1. If $G=\mathbb{Z}_3$, then X/G is a $\mathbb{Q}$-homology projective plane with 3 singular points of type $\frac{1}{3}(1,2)$ and its minimal resolution is a minimal surface of general type with $p_g=0$ and $K^2=3$.*

2. If $G=\mathbb{Z}_3^2$, then X/G is a $\mathbb{Q}$-homology projective plane with 4 singular points of type $\frac{1}{3}(1,2)$ and its minimal resolution is a minimal surface of general type with $p_g=0$ and $K^2=1$.

3. If $G=\mathbb{Z}_7$, then X/G is a $\mathbb{Q}$-homology projective plane with 3 singular points of type $\frac{1}{7}(1,5)$ and its minimal resolution is a $(2,3)$- or $(2,4)$- or $(3,3)$-elliptic surface.

4. If $G=\mathbb{Z}7:\mathbb{Z}3=G_{21}$, then X/G is a $\mathbb{Q}$-homology projective plane with 4 singular points, where three of them are of type $\frac{1}{3}(1,2)$ and one of them is of type $\frac{1}{7}(1,5)$, and its minimal resolution is a $(2,3)$- or $(2,4)$- or $(3,3)$-elliptic surface.

A fake projective plane is a nonsingular $\mathbb{Q}$-homology projective plane, hence every quotient is again a $\mathbb{Q}$-homology projective plane. An (a,b)-elliptic surface is a relatively minimal elliptic surface over $\mathbb{P}^1$ with $c_2 = 12$ having two multiple fibres of multiplicity a and b respectively. It has Kodaira dimension 1 if and only if $a \geq 2, b \geq 2, a+b \geq 5$. It is an Enriques surface iff $a = b = 2$. It is rational iff $a = 1$ or $b = 1$. All (a,b)-elliptic surfaces have $p_g = q = 0$, and by van Kampfen theorem its fundamental group is the cyclic group $\mathbb{Z}_d$ (see Dolgachev [2010]), where d is the greatest common divisor of a and b. A simply connected (a,b)-elliptic surface is called a Dolgachev surface.

Remark 2.2. The possibility of $(3,3)$-elliptic surface was further removed by the computation of Cartwright-Steger. In Cartwright and Steger [n.d.] they also computed the fundamental group of each quotient X/G, which is by the result of Armstrong [1968] isomorphic to the quotient group of the augmented fundamental group $\langle \pi_1(X), G' \rangle$ by the normal subgroup generated by elements with nonempty fixed locus on the complex 2-ball, where G' is a lifting of G onto the ball. According to their computation, $\pi_1(X/G) = \{1\}$ or $\mathbb{Z}_2$ if $G = \mathbb{Z}_7$.

2.3 Vanishing theorem for some fake projective planes. For an ample line bundle M on a fake projective plane X, M^2 is a square integer. When $M^2 \geq 9$, $H^0(X,M) \neq 0$ if and only if $M \not\cong K_X$. This follows from the Riemann-Roch and the Kodaira vanishing theorem. When $M^2 \leq 4$, $H^0(X,M)$ may not vanish, though no example of non-vanishing so far has been known. If it does not vanish, then it gives an effective curve of small degree. The non-vanishing of $H^0(X,M)$ is equivalent to the existence of certain automorphic form on the 2-ball.

Theorem 2.3 (Keum [2017]). *Let X be a fake projective plane with* $\mathrm{Aut}(X) \cong \mathbb{Z}_7 : \mathbb{Z}_3$. *Then for every $\mathbb{Z}_7$-invariant ample line bundle M with $M^2 = 4$ we have the vanishing*

$$H^0(X,M) = 0.$$

In particular, for each line bundle L with $L^2 = 1$

$$H^0(X,2L) = 0.$$

Remark 2.4.

1. A fake projective plane with $\mathrm{Aut}(X) \cong \mathbb{Z}_7 : \mathbb{Z}_3$ has only 2-torsions Cartwright and Steger [n.d.], more precisely

$$H_1(X,\mathbb{Z}) = \mathbb{Z}_2^3,\ \mathbb{Z}_2^4,\ \mathbb{Z}_2^6.$$

2. Thus K_X of such a surface has a unique cube root L_0.

3. For such a surface two ample line bundles with the same self-intersection number differ by a 2-torsion. If $M^2 = m^2$, then $M \equiv mL_0 + t$ for a 2-torsion t, hence $2M \equiv 2mL_0$ and is invariant under every automorphism.

4. The above theorem in Keum [2017] was stated only for the case $M = 2L$, but the proof used only the invariance of M under the order 7 automorphism.

5. If $M = 2L_0 + t$ is invariant under an automorphism iff so is t.

6. By Catanese and Keum [2018] the $\mathbb{Z}_7$ action on $H_1(X,\mathbb{Z})$ fixes no 2-torsion element in the case of $H_1(X,\mathbb{Z}) \cong \mathbb{Z}_2^3,\ \mathbb{Z}_2^6$, and one in the case of $H_1(X,\mathbb{Z}) \cong \mathbb{Z}_2^4$.

Theorem 2.5 (Keum [2017])**.** *Let X be a fake projective plane with* $\mathrm{Aut}(X) \cong \mathbb{Z}_3^2$*. Then for every* $\mathrm{Aut}(X)$*-invariant ample line bundle M with $M^2 = 4$ we have the vanishing*

$$H^0(X, M) = 0.$$

In particular, for the cubic root L_0 of K_X

$$H^0(X, 2L_0) = 0.$$

Remark 2.6. 1. A fake projective plane with $\mathrm{Aut}(X) \cong \mathbb{Z}_3^2$ has

$$H_1(X,\mathbb{Z}) = \mathbb{Z}_{14},\ \ \mathbb{Z}_7,\ \ \mathbb{Z}_2^2 \times \mathbb{Z}_{13}.$$

2. Thus K_X of such a surface has a unique cube root L_0.

3. For such a surface two ample line bundles with the same self-intersection number differ by a torsion.

4. The above theorem in Keum [ibid.] was stated only for the case $M = 2L_0$, but the proof used only the invariance of M under $\mathrm{Aut}(X) \cong \mathbb{Z}_3^2$.

5. If $M = 2L_0 + t$ is invariant under an automorphism iff so is t.

6. By Catanese and Keum [2018] no torsion element is $\mathbb{Z}_3^2$-invariant in the case of $H_1(X,\mathbb{Z}) \cong \mathbb{Z}_7,\ \mathbb{Z}_2^2 \times \mathbb{Z}_{13}$, and only the unique 2-torsion is $\mathbb{Z}_3^2$-invariant in the case of $H_1(X,\mathbb{Z}) \cong \mathbb{Z}_{14}$.

Both proofs used the structure of the quotients of X given in the previous subsection. The key idea of proof is that if $H^0(X, M) \neq 0$, then $\dim H^0(X, 2M) \geq 4$, contradicting the Riemann-Roch which yields $\dim H^0(X, 2M) = 3$.

2.4 Exceptionl collections of line bundles. Let $D^b(coh(W))$ denote the bounded derived category of coherent sheaves on a smooth variety W. It is a triangulated category. An object E in a triangulated category is called exceptional if $Hom(E, E[i]) = \mathbb{C}$ if $i = 0$, and $= 0$ otherwise. A sequence $E_1, \ldots, E_n$ of exceptional objects is called an exceptional sequence if $Hom(E_j, E_k[i]) = 0$ for any $j > k$, any i. When W is a smooth surface with $p_g = q = 0$, every line bundle is an exceptional object in $D^b(coh(W))$.

Let X be a fake projective plane and L be an ample line bundle with $L^2 = 1$. The three line bundles

$$2L,\ L,\ \mathcal{O}_X$$

form an exceptional sequence if and only if $H^j(X, 2L) = H^j(X, L) = 0$ for all j. Write

$$D^b(coh(X)) = \langle 2L,\ L,\ \mathcal{O}_X,\ \mathcal{A}\rangle$$

where $\mathcal{A}$ is the orthogonal complement of the admissible triangulated subcategory generated by $2L,\ L,\ \mathcal{O}_X$. Then the Hochshield homology

$$HH_*(\mathcal{A}) = 0.$$

This can be read off from the Hodge numbers. In fact, the Hochshield homology of X is the direct sum of Hodge spaces $H^{p,q}(X)$, and its total dimension is the sum of all Hodge numbers. The latter is equal to the topological Euler number $c_2(X)$, as a fake projective plane has Betti numbers $b_1(X) = b_3(X) = 0$.

The Grothendieck group $K_0(X)$ has filtration

$$K_0(X) = F^0K_0(X) \supset F^1K_0(X) \supset F^2K_0(X)$$

with

$$F^0K_0(X)/F^1K_0(X) \cong CH^0(X) \cong \mathbb{Z},$$

$$F^1K_0(X)/F^2K_0(X) \cong \mathrm{Pic}(X),$$

$$F^2K_0(X) \cong CH^2(X).$$

If the Bloch conjecture holds for X, i.e. if $CH^2(X) \cong \mathbb{Z}$, then $K_0(\mathcal{A})$ is finite.

Corollary 2.7. *Let X be a fake projective plane with $\mathrm{Aut}(X) \cong \mathbb{Z}_7 : \mathbb{Z}_3$ or $\mathbb{Z}_3^2$. Let L_0 be the unique cubic root of K_X. Then the three line bundles*

$$\mathcal{O}_X,\ -L_0,\ -2L_0$$

form an exceptional collection on X. If t is a $\mathbb{Z}_7$- or $\mathbb{Z}_3^2$-invariant torsion line bundle, then the three line bundles

$$\mathcal{O}_X,\ -(L_0+t),\ -2L_0$$

form another exceptional collection.

Such a torsion line bundle t exists for only one pair of fake projective planes in each case $\mathrm{Aut}(X) \cong \mathbb{Z}_7 : \mathbb{Z}_3$ or $\mathbb{Z}_3^2$ by Catanese and Keum [2018] (see Remark 2.4, Remark 2.6).

This is equivalent to that $H^i(X, 2L_0) = H^i(X, L_0) = H^i(X, L_0 + t) = 0$ for all i, hence follows from Theorem 2.3 and 2.5. Indeed, since L_0 is a cubic root of K_X, these vanishings are equivalent to the vanishing $H^0(X, 2L_0) = H^0(X, 2L_0 + t) = 0$. This confirms, for fake projective planes with enough automorphisms, the conjecture raised by Galkin, Katzarkov, Mellit, and Shinder [2013] that predicts the existence of an exceptional sequence of length 3 on every fake projective plane. Disjoint from our cases, Fakhruddin [2015] confirmed the conjecture for the case of three 2-adically uniformized fake projective planes found by Mumford [1979] and Ishida and Kato [1998].

2.5 Bicanonical map of fake projective planes. By Reider's theorem Reider [1988] (see Barth, Hulek, Peters, and Van de Ven [2004] for a slightly refined version) on adjoint linear systems the bicanonical system $|2K_X|$ of a ball quotient X is base point free, thus it defines a morphism.

If the ball quotient X has $\chi(X) \geq 2$, then $K_X^2 \geq 9\chi(X) \geq 10$, and since a ball quotient cannot contain a curve of geometric genus 0 or 1, the bicanonical map embeds X unless X contains a smooth genus 2 curve C with $C^2 = 0$, and $CK_X = 2$.

In the case $\chi(X) = 1$, for instance if we have a fake projective plane, we are below the Reider inequality $K_X^2 \geq 10$, and the question of the very-ampleness of the bicanonical system is interesting.

Conjecture 2.8. *For each fake projective plane its bicanonical map is an embedding into* $\mathbb{P}^9$.

Every fake projective plane X with automorphism group of order 21 cannot contain an effective curve with self-intersection 1 (Theorem 2.3), as was first proved in Keum [2013] (see Keum [2017], also Galkin, Katzarkov, Mellit, and Shinder [2015]). Thus by applying I. Reider's theorem, one sees that the bicanonical map of such a fake projective plane is an embedding into $\mathbb{P}^9$ (see also Di Brino and Di Cerbo [2018]).

Including these 3 pairs of fake projective planes, for 10 pairs the conjecture has been confirmed by Catanese and Keum [2018]. For nine pairs this follows from the vanishing result of Keum [2013, 2017], Catanese and Keum [2018]. For one pair we do not have the vanishing theorem, and the surface possesses either none or 3 curves D with $D^2 = 1$. But even in the latter case we manage to prove the very-ampleness of the bicanonical system.

2.6 Explicit equations of fake projective planes. It has long been of great interest since Mumford to find equations of a projective model of a fake projective plane.

In a recent joint work Borisov and Keum [2017, 2018] we find equations of a projective model (the bicanonical image) of a conjugate pair of fake projective planes by studying the geometry of the quotient of such surface by an order seven automorphism. The equations are given explicitly as 84 cubics in 10 variables with coefficients in the field $\mathbb{Q}[\sqrt{-7}]$. The complex conjugate equations define the bicanonical image of the complex conjugate of the surface,

This pair has the most geometric symetries among the 50 pairs, in the sense that it has the large automorphism group $G_{21} = \mathbb{Z}_7 : \mathbb{Z}_3$ and the $\mathbb{Z}_7$-quotient has a smooth model of a $(2, 4)$-elliptic surface which is not simply connected. For several pairs of fake projective planes including this pair the bicanonical map gives an embedding into the 9-dimensional projective space.

References

Valery Alexeev and Wenfei Liu (Dec. 2016). "Open surfaces of small volume". arXiv: 1612.09116 (cit. on p. 725).

Daniel Allcock and Fumiharu Kato (2017). "A fake projective plane via 2-adic uniformization with torsion". *Tohoku Math. J. (2)* 69.2, pp. 221–237. MR: 3682164.

M. A. Armstrong (1968). "The fundamental group of the orbit space of a discontinuous group". *Proc. Cambridge Philos. Soc.* 64, pp. 299–301. MR: 0221488 (cit. on p. 730).

Thierry Aubin (1976). "Équations du type Monge-Ampère sur les variétés kähleriennes compactes". *C. R. Acad. Sci. Paris Sér. A-B* 283.3, Aiii, A119–A121. MR: 0433520 (cit. on p. 728).

Wolf P. Barth, Klaus Hulek, Chris A. M. Peters, and Antonius Van de Ven (2004). *Compact complex surfaces*. Second. Vol. 4. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Springer-Verlag, Berlin, pp. xii+436. MR: 2030225 (cit. on p. 733).

Grigory Belousov (2008). "Del Pezzo surfaces with log terminal singularities". *Mat. Zametki* 83.2, pp. 170–180. MR: 2431579 (cit. on p. 719).

– (2009). "The maximal number of singular points on log del Pezzo surfaces". *J. Math. Sci. Univ. Tokyo* 16.2, pp. 231–238. MR: 2582038 (cit. on p. 719).

David Bindschadler and Lawrence Brenton (1984). "On singular 4-manifolds of the homology type of $\mathbf{C}\mathrm{P}^2$". *J. Math. Kyoto Univ.* 24.1, pp. 67–81. MR: 737825 (cit. on p. 720).

Lev Borisov and JongHae Keum (Oct. 2017). "Research announcement: equations of a fake projective plane". arXiv: 1710.04501 (cit. on p. 733).

– (Feb. 2018). "Explicit equations of a fake projective plane". arXiv: 1802.06333 (cit. on p. 733).

Lawrence Brenton (1977). "Some examples of singular compact analytic surfaces which are homotopy equivalent to the complex projective plane". *Topology* 16.4, pp. 423–433. MR: 0470253 (cit. on p. 719).

Lawrence Brenton, Daniel Drucker, and Geert C. E. Prins (1981). "Graph theoretic techniques in algebraic geometry. II. Construction of singular complex surfaces of the rational cohomology type of $\mathbf{C}P^2$". *Comment. Math. Helv.* 56.1, pp. 39–58. MR: 615614 (cit. on p. 719).

Egbert Brieskorn (1967/1968). "Rationale Singularitäten komplexer Flächen". *Invent. Math.* 4, pp. 336–358. MR: 0222084 (cit. on p. 723).

Donald I. Cartwright and Tim Steger (n.d.). "List of Fake Projective Planes" (cit. on pp. 728–730).

– (2010). "Enumeration of the 50 fake projective planes". *C. R. Math. Acad. Sci. Paris* 348.1-2, pp. 11–13. MR: 2586735 (cit. on pp. 728, 729).

Fabrizio Catanese and JongHae Keum (Jan. 2018). "The Bicanonical map of fake projective planes with an automorphism". arXiv: 1801.05291 (cit. on pp. 731, 733).

Gennaro Di Brino and Luca F. Di Cerbo (2018). "Exceptional collections and the bicanonical map of Keum's fake projective planes". *Commun. Contemp. Math.* 20.1, pp. 1650066, 13. MR: 3714832 (cit. on p. 733).

Igor Dolgachev (2010). "Algebraic surfaces with $q = p_g = 0$". In: *Algebraic surfaces*. Vol. 76. C.I.M.E. Summer Sch. Springer, Heidelberg, pp. 97–215. MR: 2757651 (cit. on p. 730).

Najmuddin Fakhruddin (2015). "Exceptional collections on 2-adically uniformized fake projective planes". *Math. Res. Lett.* 22.1, pp. 43–57. MR: 3342178 (cit. on p. 733).

Ronald Fintushel and Ronald J. Stern (1985). "Pseudofree orbifolds". *Ann. of Math. (2)* 122.2, pp. 335–364. MR: 808222 (cit. on p. 722).

– (1987). "O(2) actions on the 5-sphere". *Invent. Math.* 87.3, pp. 457–476. MR: 874031 (cit. on pp. 717, 722).

Mikio Furushima (1986). "Singular del Pezzo surfaces and analytic compactifications of 3-dimensional complex affine space $\mathbf{C}^3$". *Nagoya Math. J.* 104, pp. 1–28. MR: 868434 (cit. on p. 726).

Sergey Galkin, Ludmil Katzarkov, Anton Mellit, and Evgeny Shinder (May 2013). "Minifolds and Phantoms". arXiv: 1305.4549 (cit. on p. 733).

– (2015). "Derived categories of Keum's fake projective planes". *Adv. Math.* 278, pp. 238–253. MR: 3341791 (cit. on p. 733).

R. V. Gurjar and D.-Q. Zhang (1994). "π_1 of smooth points of a log del Pezzo surface is finite. I". *J. Math. Sci. Univ. Tokyo* 1.1, pp. 137–180. MR: 1298542.

DongSeon Hwang (n.d.). "Cascades of log del Pezzo surfaces of Picard number one". In preparation (cit. on p. 725).

DongSeon Hwang and JongHae Keum (Apr. 2009). "Algebraic Montgomery-Yang problem: the non-cyclic case". arXiv: 0904.2975 (cit. on p. 724).
– (2011a). "Algebraic Montgomery-Yang problem: the non-cyclic case". *Math. Ann.* 350.3, pp. 721–754. MR: 2805643 (cit. on pp. 723, 724).
– (2011b). "The maximum number of singular points on rational homology projective planes". *J. Algebraic Geom.* 20.3, pp. 495–523. MR: 2786664 (cit. on pp. 717, 719, 727).
– (2012). "Construction of singular rational surfaces of Picard number one with ample canonical divisor". *Proc. Amer. Math. Soc.* 140.6, pp. 1865–1879. MR: 2888175 (cit. on p. 724).
– (2013). "Algebraic Montgomery-Yang problem: the nonrational surface case". *Michigan Math. J.* 62.1, pp. 3–37. MR: 3049295 (cit. on p. 724).
– (2014). "Algebraic Montgomery-Yang problem: the log del Pezzo surface case". *J. Math. Soc. Japan* 66.4, pp. 1073–1089. MR: 3272591 (cit. on p. 724).
DongSeon Hwang, JongHae Keum, and Hisanori Ohashi (2015). "Gorenstein $\mathbb{Q}$-homology projective planes". *Sci. China Math.* 58.3, pp. 501–512. MR: 3319920 (cit. on p. 725).
Masa-Nori Ishida (1988). "An elliptic surface covered by Mumford's fake projective plane". *Tohoku Math. J. (2)* 40.3, pp. 367–396. MR: 957050.
Masa-Nori Ishida and Fumiharu Kato (1998). "The strong rigidity theorem for non-Archimedean uniformization". *Tohoku Math. J. (2)* 50.4, pp. 537–555. MR: 1653430 (cit. on pp. 728, 733).
Toshiyuki Katsura and Shigeyuki Kondō (2018). "Enriques surfaces in characteristic 2 with a finite group of automorphisms". *J. Algebraic Geom.* 27.1, pp. 173–202. MR: 3722693 (cit. on p. 726).
Toshiyuki Katsura, Shigeyuki Kondō, and Gebhard Martin (Mar. 2017). "Classification of Enriques surfaces with finite automorphism group in characteristic 2". arXiv: 1703.09609 (cit. on p. 726).
Seán Keel and James McKernan (1999). "Rational curves on quasi-projective surfaces". *Mem. Amer. Math. Soc.* 140.669, pp. viii+153. MR: 1610249 (cit. on pp. 720, 724).
JongHae Keum (2006). "A fake projective plane with an order 7 automorphism". *Topology* 45.5, pp. 919–927. MR: 2239523 (cit. on p. 728).
– (2008). "Quotients of fake projective planes". *Geom. Topol.* 12.4, pp. 2497–2515. MR: 2443971 (cit. on p. 729).
– (2010). "Projective surfaces with many nodes". In: *Algebraic geometry in East Asia—Seoul 2008*. Vol. 60. Adv. Stud. Pure Math. Math. Soc. Japan, Tokyo, pp. 245–257. MR: 2761930 (cit. on p. 718).
– (2011). "A fake projective plane constructed from an elliptic surface with multiplicities $(2,4)$". *Sci. China Math.* 54.8, pp. 1665–1678. MR: 2824965.

– (2012). "Toward a geometric construction of fake projective planes". *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* 23.2, pp. 137–155. MR: 2924897.
– (July 2013). "Every fake projective plane with an order 7 automorphism has $H^0(2L) = 0$ for any ample generator L". Manuscript circulated on July 8, 2013 (cit. on p. 733).
– (2015). "$\mathbb{Q}$-homology projective planes with nodes or cusps". In: *Algebraic geometry in east Asia—Taipei 2011*. Vol. 65. Adv. Stud. Pure Math. Math. Soc. Japan, Tokyo, pp. 143–158. MR: 3380780 (cit. on p. 718).
– (2017). "A vanishing theorem on fake projective planes with enough automorphisms". *Trans. Amer. Math. Soc.* 369.10, pp. 7067–7083. MR: 3683103 (cit. on pp. 729–731, 733).
Bruno Klingler (2003). "Sur la rigidité de certains groupes fondamentaux, l'arithméticité des réseaux hyperboliques complexes, et les "faux plans projectifs"". *Invent. Math.* 153.1, pp. 105–143. MR: 1990668 (cit. on p. 728).
János Kollár (1995). *Shafarevich maps and automorphic forms*. M. B. Porter Lectures. Princeton University Press, Princeton, NJ, pp. x+201. MR: 1341589 (cit. on p. 729).
– (2005). "Einstein metrics on five-dimensional Seifert bundles". *J. Geom. Anal.* 15.3, pp. 445–476. MR: 2190241 (cit. on p. 722).
– (2008). "Is there a topological Bogomolov-Miyaoka-Yau inequality?" *Pure Appl. Math. Q.* 4.2, Special Issue: In honor of Fedor Bogomolov. Part 1, pp. 203–236. MR: 2400877 (cit. on pp. 717, 719, 720, 722–724).
Shigeyuki Kondō (1986). "Enriques surfaces with finite automorphism groups". *Japan. J. Math. (N.S.)* 12.2, pp. 191–282. MR: 914299 (cit. on p. 726).
Vik. S. Kulikov and V. M. Kharlamov (2002). "On real structures on rigid surfaces". *Izv. Ross. Akad. Nauk Ser. Mat.* 66.1, pp. 133–152. MR: 1917540 (cit. on p. 728).
Gebhard Martin (Mar. 2017). "Enriques surfaces with finite automorphism group in positive characteristic". arXiv: 1703.08419 (cit. on p. 726).
G. Megyesi (1999). "Generalisation of the Bogomolov-Miyaoka-Yau inequality to singular surfaces". *Proc. London Math. Soc. (3)* 78.2, pp. 241–282. MR: 1665244 (cit. on p. 719).
M. Miyanishi and D.-Q. Zhang (1988). "Gorenstein log del Pezzo surfaces of rank one". *J. Algebra* 118.1, pp. 63–84. MR: 961326 (cit. on p. 726).
Yoichi Miyaoka (1984). "The maximal number of quotient singularities on surfaces with given numerical invariants". *Math. Ann.* 268.2, pp. 159–171. MR: 744605 (cit. on p. 719).
Deane Montgomery and C. T. Yang (1972). "Differentiable pseudo-free circle actions on homotopy seven spheres", 41–101. Lecture Notes in Math., Vol. 298. MR: 0362383 (cit. on pp. 719, 722).

George D. Mostow (1973). *Strong rigidity of locally symmetric spaces*. Annals of Mathematics Studies, No. 78. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, pp. v+195. MR: 0385004 (cit. on p. 728).

David Mumford (1979). "An algebraic surface with K ample, $(K^2) = 9$, $p_g = q = 0$". *Amer. J. Math.* 101.1, pp. 233–244. MR: 527834 (cit. on pp. 728, 733).

Vicente Muñoz, Juan Angel Rojo, and Aleksy Tralle (Jan. 2016). "Homology Smale-Barden manifolds with K-contact and Sasakian structures". arXiv: 1601.06136 (cit. on p. 727).

V. V. Nikulin (1979). "Integer symmetric bilinear forms and some of their geometric applications". *Izv. Akad. Nauk SSSR Ser. Mat.* 43.1, pp. 111–177, 238. MR: 525944.

Ted Petrie (1975). "Equivariant quasi-equivalence, transversality and normal cobordism", pp. 537–541. MR: 0431234 (cit. on p. 722).

Gopal Prasad and Sai-Kee Yeung (2007). "Fake projective planes". *Invent. Math.* 168, pp. 321–370. MR: 2289867 (cit. on p. 728).

– (2010). "Addendum to "Fake projective planes" Invent. Math. 168, 321–370 (2007) [MR2289867]". *Invent. Math.* 182.1, pp. 213–227. MR: 2672284 (cit. on pp. 728, 729).

Slawomir Rams and Matthias Schütt (Apr. 2014). "On Enriques surfaces with four cusps". arXiv: 1404.3924.

Igor Reider (1988). "Vector bundles of rank 2 and linear systems on algebraic surfaces". *Ann. of Math. (2)* 127.2, pp. 309–316. MR: 932299 (cit. on p. 733).

Fumio Sakai (1980). "Semistable curves on algebraic surfaces and logarithmic pluricanonical maps". *Math. Ann.* 254.2, pp. 89–120. MR: 597076 (cit. on p. 719).

Matthias Schütt (May 2015). "Moduli of Gorenstein $\mathbb{Q}$-homology projective planes". arXiv: 1505.04163 (cit. on p. 726).

– (Nov. 2016). "Q_l-cohomology projective planes from Enriques surfaces in odd characteristic". arXiv: 1611.03847 (cit. on p. 726).

– (Mar. 2017). "Q_l-cohomology projective planes and singular Enriques surfaces in characteristic two". arXiv: 1703.10441 (cit. on p. 726).

H. Seifert (1933). "Topologie Dreidimensionaler Gefaserter Räume". *Acta Math.* 60.1, pp. 147–238. MR: 1555366 (cit. on p. 721).

Giancarlo Urzúa and José I. Yáñez (Dec. 2016). "Characterization of Kollár surfaces". arXiv: 1612.01960 (cit. on p. 724).

Shing Tung Yau (1977). "Calabi's conjecture and some new results in algebraic geometry". *Proc. Nat. Acad. Sci. U.S.A.* 74.5, pp. 1798–1799. MR: 0451180 (cit. on p. 728).

Qiang Ye (2002). "On Gorenstein log del Pezzo surfaces". *Japan. J. Math. (N.S.)* 28.1, pp. 87–136. MR: 1933881 (cit. on p. 726).

De-Qi Zhang (1988). "Logarithmic del Pezzo surfaces of rank one with contractible boundaries". *Osaka J. Math.* 25.2, pp. 461–497. MR: 957874.

Received 2018-02-28.

JongHae Keum (금종해)
School of Mathematics
Korea Institute for Advanced Study
Seoul 02455
Korea
jhkeum@kias.re.kr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (737–766)

FROM CONTINUOUS RATIONAL TO REGULOUS FUNCTIONS

Wojciech Kucharz and Krzysztof Kurdyka

Abstract

Let X be an algebraic set in $\mathbb{R}^n$. Real-valued functions, defined on subsets of X, that are continuous and admit a rational representation have some remarkable properties and applications. We discuss recently obtained results on such functions, against the backdrop of previously developed theories of arc-symmetric sets, arc-analytic functions, approximation by regular maps, and algebraic vector bundles.

1 Introduction

Our purpose is to report on some new developments in real algebraic geometry, focusing on functions that have a rational representation. Let us initially consider the simplest case. A function $f : \mathbb{R}^n \to \mathbb{R}$ of class $\mathcal{C}^k$, where k is a nonnegative integer, is said to have a *rational representation* if there exist two polynomial functions p, q on $\mathbb{R}^n$ such that q is not identically 0 and $f = p/q$ on $\{q \neq 0\}$. A typical example is

(1.1) $f_k : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f_k(x, y) = \frac{x^{3+k}}{x^2 + y^2} \text{ for } (x, y) \neq (0, 0) \text{ and } f(0, 0) = 0.$$

To the best of our knowledge, Kucharz [2009] was the first paper devoted to the systematic study of such functions. This line of research was continued by several mathematicians Bilski, Kucharz, A. Valette, and G. Valette [2013], Fichou, Huisman, Mangolte, and Monnier [2016], Fichou, Monnier, and Quarez [2017], Kollár, Kucharz, and

W. Kucharz was partially supported by the National Science Center (Poland) under grant number 2014/15/ST1/00046. K. Kurdyka was partially supported by the ANR project LISA (France).
MSC2010: primary 14P05; secondary 14P99, 26C15, 57R22, 58A07.
Keywords: Real algebraic set, semialgebraic set, regular function, rational function, regulous function, arc-symmetric set, arc-analytic function, approximation, vector bundle.

Kurdyka [2018], Kollár and Nowak [2015], Kucharz [2013b, 2014a,b, 2015b, 2016a,b], Kucharz and Kurdyka [2016a,b, 2017, 2015], Kucharz and Zieliński [2018], Monnier [2018], Zieliński [2016], frequently with $k = 0$, where the functions admitting a rational representation are only continuous. Let us note that the complex case is quite different. Any continuous function from $\mathbb{C}^n$ into $\mathbb{C}$ that has a rational representation is a polynomial function.

Henceforth we work with real algebraic sets, which is equivalent to the approach adopted in Bochnak, Coste, and Roy [1998]. By a *real algebraic set* we mean an algebraic subset of $\mathbb{R}^n$ for some n. One can realize real projective d-space $\mathbb{P}^d(\mathbb{R})$ as a real algebraic set using the embedding

$$\mathbb{P}^d(\mathbb{R}) \ni (x_0 : \cdots : x_d) \mapsto \left(\frac{x_i x_j}{x_0^2 + \cdots + x_d^2} \right) \in \mathbb{R}^{(d+1)^2}. \tag{1.2}$$

Thus any algebraic subset of $\mathbb{P}^d(\mathbb{R})$ is an algebraic subset of $\mathbb{R}^{(d+1)^2}$. Consequently, many useful constructions can be performed within the class of real algebraic sets, blowing-up being an important example. One can also view any real algebraic set as the set of real points $V(\mathbb{R})$ of a quasiprojective variety V defined over $\mathbb{R}$.

Unless explicitly stated otherwise, we always assume that real algebraic sets and their subsets are endowed with the Euclidean topology, which is induced by the usual metric on $\mathbb{R}$. For a real algebraic set X, its singular locus $\operatorname{Sing}(X)$ is an algebraic, Zariski nowhere dense subset of X. We say that X is *smooth* if $\operatorname{Sing}(X)$ is empty. The following examples illustrate some phenomena that do not occur in the complex setting.

(1.3) The algebraic curve

$$C := (x^4 - 2x^2 y - y^3 = 0) \subset \mathbb{R}^2$$

is irreducible and $\operatorname{Sing}(C) = \{(0,0)\}$. Actually, C is an analytic submanifold of $\mathbb{R}^2$.

(1.4) The algebraic curve

$$C := (x^3 - x^2 - y^2 = 0) \subset \mathbb{R}^2$$

is irreducible and $\operatorname{Sing}(C) = \{(0,0)\}$. It has two connected components, the singleton $\{(0,0)\}$ and the unbounded branch $C \setminus \{(0,0)\}$.

(1.5) The algebraic curve

$$C := (x^2(x^2 - 1)(x^2 - 4) + y^2 = 0) \subset \mathbb{R}^2$$

is irreducible and $\operatorname{Sing}(C) = \{(0,0)\}$. It has three connected components, the singleton $\{(0,0)\}$ and two ovals.

(1.6) The Cartan umbrella

$$S \coloneqq (x^3 - z(x^2 + y^2) = 0) \subset \mathbb{R}^3$$

is an irreducible algebraic surface with $\operatorname{Sing}(S) = (z\text{-axis})$. The surface S is connected and $S \setminus \operatorname{Sing}(S)$ is not dense in S. Furthermore, S is not coherent when regarded as an analytic subset of $\mathbb{R}^3$.

It will be convenient to consider regular functions in a more general setting than usual. Let $X \subset \mathbb{R}^n$ be an algebraic set and let $f: W \to \mathbb{R}$ be a function defined on some subset W of X. We say that f is *regular at a point* $x \in W$ if there exist two polynomial functions p, q on $\mathbb{R}^n$ such that $q(x) \neq 0$ and $f = p/q$ on $W \cap \{q \neq 0\}$. We say that f is a *regular function* if it is regular at each point of W. For any algebraic set $Y \subset \mathbb{R}^p$, a map $\varphi = (\varphi_1, \ldots, \varphi_p): W \to Y$ is *regular* if all the components $\varphi_i: W \to \mathbb{R}$ are regular functions. These notions are independent of the algebraic embeddings $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^p$.

Any rational function R on X determines a regular function $R: X \setminus \operatorname{Pole}(R) \to \mathbb{R}$, where $\operatorname{Pole}(R)$ stands for the polar set of R.

Contents. In Section 2 we recall briefly main facts about arc-symmetric sets and arc-analytic functions. These notions, introduced 30 years ago by the second-named author, describe some rigidity phenomena (of an analytic type) of real algebraic sets. They form a background for the subsequent sections in which we present recent developments in the context of rational functions.

Section 3 contains presentation of new results on the geometry defined by regulous functions, that is, continuous functions which admit a strong version of rational representation.

In Section 4 we recall some theorems on approximation of continuous maps with values in spheres by regular maps and give new results in which approximating maps are allowed to be regulous.

In Section 5 we discuss topological, algebraic and regulous vector bundles. Regulous vector bundles have many desirable properties of algebraic vector bundles but are more flexible.

2 Arc-symmetric sets and arc-analytic functions

Arc-symmetric sets and arc-analytic functions were introduced in Kurdyka [1988]. They were further investigated and applied in Adamus and Seyedinejad [2017], Bierstone and Milman [1990], Bierstone, Milman, and Parusiński [1991], Fichou [2005], Koike and

Parusiński [2003], Kucharz [2005], Kurdyka [1991, 1994, 1999], Kurdyka and Parusiński [2007, 2012], Kurdyka and Rusek [1988], McCrory and Parusiński [2003], Parusiński [1994, 2001, 2004], Parusiński and Păunescu [2017].

2.1 Arc-symmetric sets. We say that a subset $E \subset \mathbb{R}^n$ is *arc-symmetric* if for every analytic arc $\gamma\colon (-1,1) \to \mathbb{R}^n$ with $\gamma((-1,0)) \subset E$, we have $\gamma((-1,1)) \subset E$. We are mostly interested in semialgebraic arc-symmetric sets.

Recall that a topological space is called *Noetherian* if every descending chain of its closed subsets is stationary. In particular, $\mathbb{R}^n$ with the Zariski topology is a Noetherian topological space. In Kurdyka [1988], the following is proved.

Theorem 2.1. *The semialgebraic arc-symmetric subsets of $\mathbb{R}^n$ are precisely the closed sets of a certain Noetherian topology on $\mathbb{R}^n$.*

Following Kurdyka [ibid.], we call this topology on $\mathbb{R}^n$ the $\mathfrak{AR}$ *topology*. Thus a subset of $\mathbb{R}^n$ is $\mathfrak{AR}$-closed if and only if it is semialgebraic and arc-symmetric. It follows from the curve selection lemma that each $\mathfrak{AR}$-closed subset of $\mathbb{R}^n$ is closed (in the Euclidean topology), cf. Kurdyka [ibid.]. Clearly, any connected component of an $\mathfrak{AR}$-closed subset of $\mathbb{R}^n$ is also $\mathfrak{AR}$-closed. Furthermore, any irreducible analytic component of an algebraic subset of $\mathbb{R}^n$ is $\mathfrak{AR}$-closed. However, an $\mathfrak{AR}$-closed set need not be analytic at every point.

Example 2.2. The set

$$E = \{(x,y,z) \in \mathbb{R}^3 : x^3 - z(x^2+y^2) = 0,\ x^2+y^2 \neq 0\} \cup \{(0,0,0)\}$$

(the “cloth” of the Cartan umbrella (1.6)) is $\mathfrak{AR}$-closed, but it is not analytic at the origin of $\mathbb{R}^3$.

Given a semialgebraic subset $E \subset \mathbb{R}^n$, we say that a point $x \in E$ is *regular in dimension d* if for some open neighborhood $U_x \subset \mathbb{R}^n$ of x, the intersection $E \cap U_x$ is a d-dimensional analytic submanifold of U_x. We let $\mathrm{Reg}_d(E)$ denote the locus of regular points of E in dimension d. The dimension of E, written $\dim E$, is the maximum d with $\mathrm{Reg}_d(E)$ nonempty. If V is the Zariski closure of E in $\mathbb{R}^n$, then $\dim E = \dim V$, cf. Bochnak, Coste, and Roy [1998].

By a *resolution of singularities* of a real algebraic set X we mean a proper regular map $\pi : \tilde{X} \to X$ where $\tilde{X}$ is a smooth real algebraic set and π is birational.

The following is the key result of Kurdyka [1988].

Theorem 2.3. *Let $X \subset \mathbb{R}^n$ be a d-dimensional real algebraic set and let $E \subset \mathbb{R}^n$ be an $\mathfrak{AR}$-closed irreducible subset with $E \subset X$ and* $\dim E = d$. *If $\pi : \tilde{X} \to X$ is a resolution of singularities of X, then there exists a unique connected component $\tilde{E}$ of $\tilde{X}$ such that $\pi(\tilde{E})$ is the closure (in the Euclidean topology) of* $\mathrm{Reg}_d(E)$.

This is illustrated by an example below.

Example 2.4. The real cubic $C := (x^3 - x - y^2 = 0) \subset \mathbb{R}^2$ is smooth and irreducible. It has two connected components, C_1 which is compact and C_2 which is noncompact. Consider the cone $X := (x^3 - xz^2 - y^2 z = 0) \subset \mathbb{R}^3$ over C. Note that X is irreducible and $\operatorname{Sing}(X) = \{(0,0,0)\}$. Clearly,

$$\pi\colon \tilde{X} := C \times \mathbb{R} \to X, \quad (x, y, z) \mapsto (xz, yz, z)$$

is a resolution of singularities of X. The connected components $C_1 \times \mathbb{R}$ and $C_2 \times \mathbb{R}$ of $\tilde{X}$ correspond via π to the $\mathfrak{AR}$-irreducible components of X.

The notion of arc-symmetric set turns out to be related to a notion introduced by Nash in his celebrated paper Nash [1952]. We adapt his definition to the case of $\mathfrak{AR}$-closed sets.

Definition 2.5. Let E be an $\mathfrak{AR}$-closed subset of $\mathbb{R}^n$. We say that a subset $S \subset E$ is a (Nash) *sheet* of E if the following conditions are satisfied:

(i) for any two points x_0, x_1 in S there exists an analytic arc $\gamma\colon [0,1] \to \mathbb{R}^n$ with $\gamma(0) = x_0$, $\gamma(1) = x_1$, and $\gamma([0,1]) \subset S$;

(ii) S is maximal in the class of subsets satisfying the condition (i);

(iii) the interior of S in X is nonempty.

The following result of Kurdyka [1988] gives a positive and precise answer to Nash's conjecture on sheets of an algebraic set Nash [1952].

Theorem 2.6. *Let X be an algebraic subset or more generally an $\mathfrak{AR}$-closed subset of $\mathbb{R}^n$. Then:*

(i) *There are finitely many sheets in X.*

(ii) *Each sheet in X is semialgebraic and closed (in the Euclidean topology).*

(iii) *X is the union of its sheets.*

The proof of this theorem is based on Theorem 2.3 and the notion of *immersed component* of an $\mathfrak{AR}$-closed set. An immersed component of X is an $\mathfrak{AR}$-irreducible subset of X with nonempty interior in X. In general, X may have more immersed components than $\mathfrak{AR}$-irreducible components. For instance, the Whitney umbrella $(xy^2 - z^2 = 0) \subset \mathbb{R}^3$ is $\mathfrak{AR}$-irreducible, but it has two immersed components.

Compact $\mathfrak{AR}$-closed sets share with compact real algebraic sets all known local and global topological properties. In particular, each compact $\mathfrak{AR}$-closed set carries the mod 2 fundamental class. It is conjectured that any compact $\mathfrak{AR}$-closed set is semialgebraically homeomorphic to a real algebraic set.

Recall that a Nash manifold $X \subset \mathbb{R}^n$ is an analytic submanifold which is a semialgebraic set. Building on Thom's representability theorem Thom [1954] and Theorem 2.3, the following was established in Kucharz [2005].

Theorem 2.7. *Let $X \subset \mathbb{R}^n$ be a compact Nash manifold, and d an integer satisfying $0 \leq d \leq \dim X$. Then each homology class in $H_d(X;\mathbb{Z}/2)$ can be represented by an $\mathfrak{AR}$-closed subset of $\mathbb{R}^n$, contained in X.*

Now we recall a result of Kurdyka and Rusek [1988] which was motivated by the problem of surjectivity of injective selfmaps.

Theorem 2.8. *Let $X \subset \mathbb{R}^n$ be an $\mathfrak{AR}$-closed subset of dimension d, with $0 \leq d \leq n-1$. Then the homotopy group $\pi_{n-d-1}(\mathbb{R}^n \setminus X)$ is nontrivial.*

As demonstrated in Kurdyka and Rusek [ibid.], Theorem 2.8 implies the following result of Białynicki-Birula and Rosenlicht [1962].

Theorem 2.9. *Any injective polynomial map from $\mathbb{R}^n$ into itself is surjective.*

One should mention that Theorem 2.9, with $n = 2$, was established earlier by Newman [1960]. Ax [1969] proved that any injective regular map of a complex algebraic variety into itself is surjective. Ax's proof is based on the Lefschetz principle and a reduction to the finite field case. By extending the idea of Białynicki-Birula and Rosenlicht [1962], Borel [1969] gave a topological proof of Ax's theorem that works also for injective regular maps of a smooth real algebraic set into itself. Finally, combining Borel's argument with the geometry of $\mathfrak{AR}$-closed sets, the second-named author proved in Kurdyka [1999] the following.

Theorem 2.10. *Let X be a real algebraic set (possibly singular) and let $f : X \to X$ be an injective regular map. Then f is surjective.*

In fact, there is a more general version of Theorem 2.10 due to Parusiński [2004], cf. also Kurdyka and Parusiński [2007].

2.2 Arc-analytic maps. Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^p$ be some subsets. A map $f : X \to Y$ is said to be *arc-analytic* if for every analytic arc $\gamma : (-1,1) \to \mathbb{R}^n$ with $\gamma((-1,1)) \subset X$, the composite $f \circ \gamma : (-1,1) \to \mathbb{R}^p$ is an analytic map. We are mostly interested in the case where X and Y are $\mathfrak{AR}$-closed, and f is semialgebraic.

The function $f_k\colon \mathbb{R}^2 \to \mathbb{R}$ in (1.1) is arc-analytic and of class $\mathcal{C}^k$, but it is not of class $\mathcal{C}^{k+1}$. The following fact is recorded in Kurdyka [1988].

Proposition 2.11. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^p$ be $\mathcal{AR}$-closed subsets, and let $f\colon X \to Y$ be a semialgebraic arc-analytic map. Then:*

(i) *The graph of f is an $\mathcal{AR}$-closed subset of $\mathbb{R}^n \times \mathbb{R}^p$.*

(ii) *If $Z \subset Y$ is an $\mathcal{AR}$-closed set, then so is $f^{-1}(Z)$.*

(iii) *f is continuous (in the Euclidean topology).*

Arc-analytic functions do not have nice properties without some additional assumptions. For instance, an arc-analytic function on $\mathbb{R}^n$ need not be subanalytic Kurdyka [1991] or continuous Bierstone, Milman, and Parusiński [1991], and even for $n = 2$ it may have a nondiscrete singular set Kurdyka [1994].

In complex algebraic geometry, the image of an algebraic set by a proper regular map is again an algebraic set. This is trivially false in the real case; consider $\mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$. There is also a more interesting example. Let

$$X := (x - y^2 - 1 = 0) \subset \mathbb{R}^2 \text{ and } Y := (x^3 - x^2 - y^2 = 0) \subset \mathbb{R}^2.$$

Then $f\colon X \to Y$, $(x, y) \mapsto (x, xy)$ is an injective, proper regular map. However, $f(X)$ is not an algebraic set. Therefore the following embedding theorem of Kurdyka [1988] is of interest.

Theorem 2.12. *Let $X \subset \mathbb{R}^n$ be an $\mathcal{AR}$-closed subset and let $f\colon X \to \mathbb{R}^p$ be a semialgebraic arc-analytic map that is injective and proper. Then $f(X) \subset \mathbb{R}^p$ is an $\mathcal{AR}$-closed subset.*

Given an $\mathcal{AR}$-closed subset $X \subset \mathbb{R}^n$, we denote by $\mathcal{A}_a(X)$ the ring of semialgebraic arc-analytic functions on X. According to Kurdyka [ibid.], the ring $\mathcal{A}_a(X)$ is not Noetherian if $\dim X \geq 2$. However, any ascending chain of prime ideals of $\mathcal{A}_a(X)$ is stationary. Furthermore, by Kurdyka [ibid.], there exists a function $f \in \mathcal{A}_a(\mathbb{R}^n)$ such that $X \subset f^{-1}(0)$ and $\dim(f^{-1}(0) \setminus X) < \dim X$. This latter result has been recently strengthened by Adamus and Seyedinejad [2017], who proved that actually $X = f^{-1}(0)$ for some $f \in \mathcal{A}_a(\mathbb{R}^n)$. This enabled them to obtain the Nullstellensatz for the ring $\mathcal{A}_a(X)$, generalizing thereby the weak Nullstellensatz of Kurdyka [1988].

2.3 Blow-Nash and blow-analytic functions.

Let X be a smooth real algebraic set. A *Nash function* on X is an analytic function which is semialgebraic. A function on X is said to be *blow-Nash* if it becomes a Nash function after composing with a finite sequence of blowups with smooth nowhere dense centers. It was conjectured by K. Kurdyka (1987) that a function is blow-Nash if and only if it is arc-analytic and semialgebraic. The first proof of this conjecture was published by Bierstone and Milman [1990]. They developed techniques which later turned out to be useful in their approach to the resolution of singularities Bierstone and Milman [1997]. There is also a second proof due to Parusiński [1994]. It is based on the rectilinearization theorem for subanalytic functions Parusiński [ibid.], which is a prototype of the preparation theorem for subanalytic functions, cf. for example Parusiński [2001].

Less is known on arc-analytic functions which are subanalytic. Any such function is continuous and can be made analytic after composing with finitely many local blowups with smooth centers, cf. Bierstone and Milman [1990] and Parusiński [1994]. It is not known whether one can use global blowups, that is, whether arc-analytic subanalytic functions coincide with *blow-analytic* functions of T. C. Kuo [1985], cf. also Fukui, Koike, and T.-C. Kuo [1998]. In Kurdyka and Parusiński [2012] it is proved that the locus of nonanalyticity of an arc-analytic subanalytic function is arc-symmetric and subanalytic. Another result of Kurdyka and Parusiński [ibid.] asserts that in the blow-Nash case, the centers of blowups can be chosen in the locus of nonanalyticity.

2.4 Some applications. Recently arc-symmetric sets were used in the construction of new invariants in the singularity theory. These invariants include the virtual Betti numbers of real algebraic sets McCrory and Parusiński [2003] and arc-symmetric sets Fichou [2005]. Other invariants, analogous to the zeta function of Denef and Loeser, proved to be useful in the classification of germs of functions with respect to blow-analytic and blow-Nash equivalence, cf. Koike and Parusiński [2003] and Fukui, Koike, and T.-C. Kuo [1998]. Arc-analytic homeomorphisms were recently used in Parusiński and Păunescu [2017] to construct nice trivializations in the stratification theory.

3 Regulous functions

3.1 Functions regular on smooth algebraic arcs. All results presented in this subsection come from our joint paper Kollár, Kucharz, and Kurdyka [2018].

Let X be a real algebraic set. A subset $A \subset X$ is called a *smooth algebraic arc* if its Zariski closure C is an irreducible algebraic curve, $A \subset C \setminus \operatorname{Sing}(C)$, and A is homeomorphic to $\mathbb{R}$.

An open subset $U \subset X$ is said to be *smooth* if it is contained in $X \setminus \operatorname{Sing}(X)$.

Theorem 3.1. *Let X be a real algebraic set and let $f : U \to \mathbb{R}$ be a function defined on a connected smooth open subset $U \subset X$. Assume that the restriction of f is regular on* *each smooth algebraic arc contained in* U. *Then there exists a rational function* R *on* X *such that* $P := U \cap \operatorname{Pole}(R)$ *has codimension at least* 2 *and* $f|_{U\setminus P} = R|_{U\setminus P}$.

There are two main steps in the proof of Theorem 3.1. Assuming that f is a semialgebraic function (so U is a semialgebraic set), one first obtains a local variant of the assertion by means of Bertini's theorem, and then extends it along smooth algebraic arcs. The general case is reduced to the semialgebraic one via some subtle Hartogs-like results on analytic functions due to Błocki [1992] and Siciak [1990].

A function regular on smooth algebraic arcs need not be continuous.

Example 3.2. The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \frac{x^8 + y(x^2 - y^3)^2}{x^{10} + (x^2 - y^3)^2} \text{ for } (x,y) \neq (0,0) \text{ and } f(0,0) = 0$$

is regular on each smooth algebraic arc in $\mathbb{R}^2$, but it is not locally bounded on the curve $x^2 - y^3 = 0$.

Let $X = X_1 \times \cdots \times X_n$ be the product of real algebraic sets and let $\pi_i : X \to X_i$ be the projection on the ith factor. A subset $K \subset X$ is said to be *parallel to the* i*th factor of* X if $\pi_j(K)$ consists of one point for each $j \neq i$.

Theorem 3.3. *Let* $X = X_1 \times \cdots \times X_n$ *be the product of real algebraic sets and let* $f : U \to \mathbb{R}$ *be a function defined on a connected smooth open subset* $U \subset X$. *Assume that the restriction of* f *is regular on each smooth algebraic arc contained in* U *and parallel to one of the factors of* X. *Then there exists a rational function* R *on* X *such that* $P := U \cap \operatorname{Pole}(R)$ *has codimension at least* 2 *and* $f|_{U\setminus P} = R|_{U\setminus P}$.

Theorem 3.3, for $n = 1$, coincides with Theorem 3.1. The general case is proved by induction on n, but a detailed argument is fairly long.

As a direct consequence, we get the following.

Corollary 3.4. *Let* $f : U \to \mathbb{R}$ *be a function defined on a connected open subset* $U \subset \mathbb{R}^n$. *Assume that the restriction of* f *is regular on each open interval contained in* U *and parallel to one of the coordinate axes. Then there exists a rational function* R *on* $\mathbb{R}^n$ *such that* $P := U \cap \operatorname{Pole}(R)$ *has codimension at least* 2 *and* $f|_{U\setminus P} = R|_{U\setminus P}$.

Results similar to Corollary 3.4 have been known earlier, but they were obtained under the restrictive assumption that f is an analytic function on U, cf. Bochner and Martin [1948].

3.2 Introducing regulous functions. Let X be a real algebraic set, $f : W \to \mathbb{R}$ a function defined on some subset $W \subset X$, and Y the Zariski closure of W in X.

Definition 3.5. A rational function R on Y is said to be a *rational representation* of f if there exists a Zariski open dense subset $Y^0 \subset Y \setminus \operatorname{Pole}(R)$ such that $f|_{W \cap Y^0} = R|_{W \cap Y^0}$.

While the definition makes sense for an arbitrary subset W, it is sensible only if W contains a sufficiently large portion of Y. The key examples of interest are open subsets and semialgebraic subsets, with $W = X$ being the most important case.

One readily checks that the following conditions are equivalent:

(3.6) For every algebraic subset $Z \subset X$ the restriction $f|_{W \cap Z}$ has a rational representation.

(3.7) There exists a sequence of algebraic subsets

$$X = X_0 \supset X_1 \supset \cdots \supset X_{m+1} = \varnothing$$

such that the restriction of f is regular on $W \cap (X_i \setminus X_{i+1})$ for $i = 0, \dots, m$.

(3.8) There exists a finite stratification $\mathcal{S}$ of X, with Zariski locally closed strata, such that the restriction of f is regular on $W \cap S$ for every $S \in \mathcal{S}$.

Definition 3.9. We say that f is a *regulous function* if it is continuous and the equivalent conditions (3.6), (3.7), (3.8) are satisfied.

In some papers, regulous functions are called *hereditarily rational* Kollár, Kucharz, and Kurdyka [2018] and Kollár and Nowak [2015] or *stratified-regular* Kucharz [2015b], Kucharz and Kurdyka [2016b, 2015], and Zieliński [2016]. The short name "regulous", derived from "regular" and "continuous", was introduced in Fichou, Huisman, Mangolte, and Monnier [2016]. A continuous function that has a rational representation is often called simply a *continuous rational function* Kollár, Kucharz, and Kurdyka [2018], Kollár and Nowak [2015], Kucharz [2009, 2013b, 2014b,a, 2016a], and Kucharz and Kurdyka [2016a, 2017].

Evidently, any regulous function is continuous and has a rational representation. The converse holds in an important special case.

Proposition 3.10. *Let X be a real algebraic set and let $W \subset X$ be a smooth open subset. For a function $f : W \to \mathbb{R}$, the following conditions are equivalent:*

(a) *f is regulous.*

(b) *f is continuous and has a rational representation.*

The nontrivial implication (b)⇒(a) is proved in Kollár and Nowak [2015]. Suppose that (b) holds, and let R be a rational representation of f. Since f is continuous and W

is smooth, one gets $f|_{W\setminus P} = R|_{W\setminus P}$, where $P = W \cap \operatorname{Pole}(R)$. Furthermore, it is not hard to see that P has codimension at least 2. Finally, condition (3.6) can be verified by induction on $\operatorname{codim} Z$.

As demonstrated in Kollár and Nowak [ibid.] and recalled below, the smoothness assumption in Proposition 3.10 cannot be dropped.

Example 3.11. Consider the algebraic surface

$$S := (x^3 - (1+z^2)y^3 = 0) \subset \mathbb{R}^3$$

and the function $f : S \to \mathbb{R}$ defined by $f(x,y,z) = (1+z^2)^{1/3}$. Note that $\operatorname{Sing}(S) =$ z-axis and $f(x,y,z) = x/y$ on $S \setminus (z\text{-axis})$. In particular, f is continuous and has a rational representation. However, f is not regulous since $f|_{z\text{-axis}}$ does not have a rational representation. It is also interesting that S is an analytic submanifold of $\mathbb{R}^3$.

The main result of Kollár and Nowak [ibid.] can be stated as follows.

Theorem 3.12. *Let X be a smooth real algebraic set and let $f : W \to \mathbb{R}$ be a function defined on an algebraic subset $W \subset X$. Then the following conditions are equivalent:*

(a) *f is regulous.*

(b) *$f = \tilde{f}|_W$, where $\tilde{f} : X \to \mathbb{R}$ is a continuous function that has a rational representation.*

The proof of (a)⇒(b) by induction on $\dim W$ is tricky. Roughly speaking, one first finds a function on X that extends f and has a rational representation. However, such an extension may not be continuous and has to be corrected. This is achieved by analyzing liftings of functions to the blowup of X at a suitably chosen ideal. The argument relies on a version of the Łojasiewicz inequality given in Bochnak, Coste, and Roy [1998, Theorem 2.6.6].

The implication (b)⇒(a) follows from Proposition 3.10.

As it was noted on various occasions (see for example Kucharz [2009, p. 528] or Fichou, Huisman, Mangolte, and Monnier [2016, Théorème 3.11]), Hironaka's theorem on resolution of indeterminacy points Hironaka [1964] implies immediately the following.

Proposition 3.13. *Let X be a smooth real algebraic set. For a function $f : X \to \mathbb{R}$, the following conditions are equivalent:*

(a) *f is continuous and has a rational representation.*

(b) *There exists a regular map $\pi : X' \to X$, which is the composite of a finite sequence of blowups with smooth Zariski nowhere dense centers, such that the function $f \circ \pi : X' \to \mathbb{R}$ is regular.*

Fefferman and Kollár [2013] study the following problem. Consider a linear equation

$$f_1 y_1 + \cdots + f_r y_r = g,$$

where g and the f_i are regular (or polynomial) functions on $\mathbb{R}^n$. Assume that it admits a solution where the y_i are continuous functions on $\mathbb{R}^n$. Then, according to Fefferman and Kollár [ibid., Section 2], it has also a continuous semialgebraic solution. One could hope to prove that it has a regulous solution. This is indeed the case for $n = 2$ Kucharz and Kurdyka [2017, Corollary 1.7], but fails for any $n \geq 3$ Kollár and Nowak [2015, Example 6]. It would be interesting to decide which linear equations have regulous solutions. Of course, the problem can be considered in a more general setting, replacing $\mathbb{R}^n$ by a real algebraic set.

3.3 Curve-regulous and arc-regulous functions.

All results discussed in this subsection come from our joint paper Kollár, Kucharz, and Kurdyka [2018], where it is proved that regulous functions can be characterized by restrictions to algebraic curves or algebraic arcs.

Definition 3.14. Let X be a real algebraic set and let $f : W \to \mathbb{R}$ be a function defined on some subset $W \subset X$.

We say that f is *regulous on algebraic curves* or *curve-regulous* for short if for every irreducible algebraic curve $C \subset X$ the restriction $f|_{W \cap C}$ is regulous (equivalently, $f|_{W \cap C}$ is continuous and has a rational representation).

Furthermore, we say that f is *regulous on algebraic arcs* or *arc-regulous* for short if for every irreducible algebraic curve $C \subset X$ and every point $x \in W \cap C$ there exists an open neighborhood $U_x \subset W$ of x such that the restriction $f|_{U_x \cap C}$ is regulous (equivalently, $f|_{U_x \cap C}$ is continuous and has a rational representation).

Obviously, any curve-regulous function is arc-regulous. The converse does not hold for a rather obvious reason. For instance, consider the hyperbola $H \subset \mathbb{R}^2$ defined by $xy - 1 = 0$. Any function on H that is constant on each connected component of H is arc-regulous, but it must be constant to be regulous.

In Kollár, Kucharz, and Kurdyka [ibid.], curve-regulous (resp. arc-regulous) functions are called *curve-rational* (resp. *arc-rational*).

Our main result on curve-regulous functions is the following.

Theorem 3.15. *Let X be a real algebraic set and let $W \subset X$ be a subset that is either open or semialgebraic. For a function $f : W \to \mathbb{R}$, the following conditions are equivalent:*

(a) *f is regulous.*

(b) *f is curve-regulous.*

The corresponding result for arc-regulous functions takes the following form.

Theorem 3.16. *Let X be a real algebraic set and let $W \subset X$ be a connected smooth open subset. For a function $f : W \to \mathbb{R}$, the following conditions are equivalent:*

(a) *f is regulous.*

(b) *f is arc-regulous.*

The crucial ingredient in the proofs of Theorems 3.15 and 3.16 is Theorem 3.1. In both cases, only the implication (b)⇒(a) is not obvious. It is essential that testing curves and arcs are allowed to have singularities.

The main properties of arc-regulous functions on semialgebraic sets can be summarized as follows.

Theorem 3.17. *Let X be a real algebraic set and let $f : W \to \mathbb{R}$ be an arc-regulous function defined on a semialgebraic subset $W \subset X$. Then f is continuous and there exists a sequence of semialgebraic sets*

$$W = W_0 \supset W_1 \supset \cdots \supset W_{m+1} = \varnothing,$$

which are closed in W, such that the restriction of f is a regular function on each connected component of $W_i \setminus W_{i+1}$ for $i = 0, \ldots, m$. In particular, f is a semialgebraic function.

We also establish a connection between arc-regulous functions and, discussed in Section 2, arc-analytic functions.

Theorem 3.18. *Let X be a real algebraic set and let $f : W \to \mathbb{R}$ be an arc-regulous function defined on an open subset $W \subset X$. Then f is continuous and arc-analytic.*

In Kollár, Kucharz, and Kurdyka [ibid.] there are several other related results.

3.4 Constructible topology and k-regulous functions. We consider regulous functions of class $\mathcal{C}^k$.

Definition 3.19. Let X be a smooth real algebraic set and let $f : U \to \mathbb{R}$ be a function defined on an open subset $U \subset X$.

We say that f is a *k-regulous* function, where k is a nonnegative integer, if it is of class $\mathcal{C}^k$ and regulous; or equivalently, by Proposition 3.10, if it is of class $\mathcal{C}^k$ and has a rational representation.

The set $\mathcal{R}^k(U)$ of all k-regulous functions on U forms a ring. An example of a k-regulous function on $\mathbb{R}^2$ is provided by (1.1).

A function on U which is of class $\mathcal{C}^\infty$ and regulous is actually regular, cf. Kucharz [2009]. Therefore one gains no new insight by considering such functions.

All results discussed in the remainder of this subsection come from the paper of Fichou, Huisman, Mangolte, and Monnier [2016], where they are stated for functions defined on $\mathbb{R}^n$. The ring $\mathcal{R}^k(\mathbb{R}^n)$ is not Noetherian if $n \geq 2$. Nevertheless it has some remarkable properties.

Given a collection F of real-valued functions on $\mathbb{R}^n$, we set

$$Z(F) := \{x \in \mathbb{R}^n : f(x) = 0 \text{ for all } f \in F\}$$

and write $Z(f)$ for $Z(F)$ if $F = \{f\}$.

The following is a variant of the classical Nullstellensatz for the ring $\mathcal{R}^k(\mathbb{R}^n)$.

Theorem 3.20. *Let I be an ideal of the ring $\mathcal{R}^k(\mathbb{R}^n)$. If a function f in $\mathcal{R}^k(\mathbb{R}^n)$ vanishes on $Z(I)$, then f^m belongs to I for some positive integer m.*

Recall that the Nullstellensatz for the ring of polynomial or regular functions on $\mathbb{R}^n$ requires an entirely different formulation, cf. Bochnak, Coste, and Roy [1998].

The subsets of $\mathbb{R}^n$ of the form $Z(I)$ for some ideal I of $\mathcal{R}^k(\mathbb{R}^n)$ can be characterized in terms of constructible sets. A subset of $\mathbb{R}^n$ is said to be *constructible* if it belongs to the Boolean algebra generated by the algebraic subsets of $\mathbb{R}^n$; or equivalently if it is a finite union of Zariski locally closed subsets of $\mathbb{R}^n$.

Theorem 3.21. *For a subset $E \subset \mathbb{R}^n$, the following conditions are equivalent:*

(a) *$E = Z(I)$ for some ideal I of $\mathcal{R}^k(\mathbb{R}^n)$.*

(b) *$E = Z(f)$ for some function f in $\mathcal{R}^k(\mathbb{R}^n)$.*

(c) *E is closed and constructible.*

Theorem 3.21 can be illustrated as follows.

Example 3.22. Consider the Cartan umbrella $S \subset \mathbb{R}^3$ defined in (1.6), and let E be the closure of $S \setminus (z\text{-axis})$. It is clear that E is a closed constructible set. Moreover, $E = Z(f)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the regulous function defined by

$$f(x, y, z) = z - \frac{x^3}{x^2 + y^2} \text{ on } \mathbb{R}^3 \setminus (z\text{-axis}) \text{ and } f(x, y, z) = z \text{ on the } z\text{-axis}.$$

The collection of all closed constructible subsets of $\mathbb{R}^n$ forms the family of closed subsets for a Noetherian topology on $\mathbb{R}^n$, called the *constructible topology*. Any subset of $\mathbb{R}^n$ which is constructible-closed is actually $\mathcal{AR}$-closed. The converse does not hold if $n \geq 2$.

In what follows we consider $\mathbb{R}^n$ endowed with the constructible topology. The assignment $\mathcal{R}^k : U \mapsto \mathcal{R}^k(U)$, where U runs through the open subsets of $\mathbb{R}^n$, is a sheaf of rings on $\mathbb{R}^n$, and $(\mathbb{R}^n, \mathcal{R}^k)$ is a locally ringed space. Sheaves of $\mathcal{R}^k$-modules on $\mathbb{R}^n$ are called *k-regulous sheaves*.

It follows from Theorem 3.20 that the ringed space $(\mathbb{R}^n, \mathcal{R}^k)$ carries essentially the same information as the affine scheme $\operatorname{Spec}(\mathcal{R}^k(\mathbb{R}^n))$. In particular, Cartan's theorems A and B are available for k-regulous sheaves.

Theorem 3.23. *For any quasi-coherent k-regulous sheaf $\mathcal{F}$ on $\mathbb{R}^n$, the following hold:*

(A) *$\mathcal{F}$ is generated by global sections.*

(B) *$H^i(\mathbb{R}^n, \mathcal{F}) = 0$ for $i \geq 1$.*

As is well-known, Cartan's theorems A and B fail for coherent algebraic sheaves on $\mathbb{R}^n$.

Let $V \subset \mathbb{R}^n$ be a constructible-closed subset. The sheaf $\mathcal{J}_V \subset \mathcal{R}^k$ of ideals of k-regulous functions vanishing on V is a quasi-coherent k-regulous sheaf on $\mathbb{R}^n$, and the quotient sheaf $\mathcal{R}^k/\mathcal{J}_V$ has support V. Endow V with the induced (constructible) topology, and let $\mathcal{R}^k_V$ be the restriction of the sheaf $\mathcal{R}^k_V/\mathcal{J}_V$ to V. The locally ringed space $(V, \mathcal{R}^k_V)$ is called a *closed k-regulous subvariety of* $(\mathbb{R}^n, \mathcal{R}^k)$. One can consider k-regulous sheaves on V, which are just sheaves of $\mathcal{R}^k_V$-modules. By a standard argument, Theorem 3.23 implies that Cartan's theorems A and B hold also for quasi-coherent k-regulous sheaves on V.

An *affine k-regulous variety* is a locally ringed space isomorphic to a closed k-regulous subvariety of $\mathbb{R}^n$ for some n. An *abstract k-regulous variety* can be defined in the standard way. The geometry of k-regulous varieties is to be developed.

4 Homotopy and approximation

In this section we discuss some homotopy and approximation results in the framework of real algebraic geometry, focusing on maps with values in the unit p-sphere

$$\mathbb{S}^p = (u_0^2 + \cdots + u_p^p - 1 = 0) \subset \mathbb{R}^{p+1}.$$

Approximation of continuous maps means approximation in the compact-open topology.

4.1 Approximation by regular maps. The theory of regular maps between real algebraic sets was developed by J. Bochnak and the first-named author Bochnak and Kucharz

[1987a,b, 1988, 1989b, 1991b, 1993, 1999, 2007] who joined forces with R. Silhol working on Bochnak, Kucharz, and Silhol [1997]. Regular maps are studied also in Ghiloni [2006, 2007], Joglar-Prieto and Mangolte [2004], Kucharz [2010, 2013a, 2015a], Loday [1973], Ozan [1995], Peng and Tang [1999], Turiel [2007], and Wood [1968]. We make no attempt to survey this theory, but give instead a sample of results that motivated later work described in the next subsection.

Problem 4.1. Let X be a compact real algebraic set. For a continuous map

$$f : X \to \mathbb{S}^p,$$

consider the following questions:

(i) Is f homotopic to a regular map?

(ii) Can f be approximated by regular maps?

It is expected that questions (i) and (ii) are equivalent, however, the proof is available only for special values of p, cf. Bochnak and Kucharz [1987a].

Theorem 4.2. *Let X be a compact real algebraic set. For a continuous map*

$$f : X \to \mathbb{S}^p,$$

where $p \in \{1, 2, 4\}$, the following conditions are equivalent:

(a) *f is homotopic to a regular map.*

(b) *f can be approximated by regular maps.*

Basic topological properties of regular maps between unit spheres still remain mysterious.

Conjecture 4.3. *For any pair (n, p) of positive integers, the following assertions hold:*

(i) *Each continuous map from $\mathbb{S}^n$ into $\mathbb{S}^p$ is homotopic to a regular map.*

(ii) *Each continuous map from $\mathbb{S}^n$ into $\mathbb{S}^p$ can be approximated by regular maps.*

Conjecture 4.3 (i) is known to be true in several cases Bochnak, Coste, and Roy [1998], Bochnak and Kucharz [1987a,b], Peng and Tang [1999], Turiel [2007], and Wood [1968]; for example if $n = p$ or $(n, p) = (2q + 14, 2q + 1)$ with $q \geq 7$.

Conjecture 4.3 (ii) holds if either $n < p$ (trivial) or $p \in \{1, 2, 4\}$ Bochnak and Kucharz [1987a]. Nothing is known for other pairs (n, p).

However, a complete solution to Problem 4.1 is known in several cases. The simplest one, noted in Bochnak and Kucharz [ibid.], is the following.

Proposition 4.4. *Let X be a compact smooth real algebraic curve. Then each continuous map from X into $\mathbb{S}^1$ can be approximated by regular maps.*

Going beyond curves is a lot harder. Nevertheless, it can happen also in higher dimension that the behavior of regular maps is determined entirely by the topology of the real algebraic sets involved.

Consider a compact $\mathcal{C}^\infty$ manifold M. A smooth real algebraic set diffeomorphic to M is called an *algebraic model* of M. By the Nash–Tognoli theorem Nash [1952] and Tognoli [1973], M has algebraic models. Actually, according to Bochnak and Kucharz [1991a], there exists an uncountable family of pairwise birationally nonequivalent algebraic models of M, provided that $\dim M \geq 1$. If $\dim M \leq 2$, the existence of algebraic models of M follows easily from the well-known classification of such manifolds. As X runs through the class of all algebraic models of M, the topological properties of regular maps from X into $\mathbb{S}^p$ may vary; this phenomenon is extensively investigated in Bochnak and Kucharz [1987b, 1988, 1989b, 1990, 1993] and Kucharz [2010].

A detailed study of regular maps into $\mathbb{S}^1$ is contained in Bochnak and Kucharz [1989b] where in particular the following result is proved.

Theorem 4.5. *Let M be a compact $\mathcal{C}^\infty$ manifold. Then there exists an algebraic model X of M such that each continuous map from X into $\mathbb{S}^1$ can be approximated by regular maps.*

For simplicity, we state the next result of Bochnak and Kucharz [ibid.] only for surfaces.

Theorem 4.6. *Let M be a connected, compact $\mathcal{C}^\infty$ surface. Then the following conditions are equivalent:*

(a) *For any algebraic model X of M, each continuous map from X into $\mathbb{S}^1$ can be approximated by regular maps.*

(b) *M is homeomorphic to the unit 2-sphere or the real projective plane or the Klein bottle.*

In Bochnak and Kucharz [1987b], one finds the following.

Theorem 4.7. *Let M be a compact $\mathcal{C}^\infty$ manifold of dimension p. Then there exists an algebraic model X of M such that each continuous map from X into $\mathbb{S}^p$ is homotopic to a regular map.*

Theorem 4.7, for $p = 1$, is of course weaker than Proposition 4.4. The cases $p = 2$ and $p = 4$ are of particular interest in view of Theorem 4.2.

Numerous results on algebraic models and regular maps into even-dimensional spheres are included in Bochnak and Kucharz [1988, 1990, 1993] and Kucharz [2010]. The following comes from Bochnak and Kucharz [1988].

Theorem 4.8. *Let M be a connected, compact $\mathcal{C}^\infty$ surface. Then the following conditions are equivalent:*

(a) *For any algebraic model X of M, each continuous map from X into $\mathbb{S}^2$ can be approximated by regular maps.*

(b) *M is nonorientable of odd genus.*

The true complexity of Problem 4.1 becomes apparent for surfaces of other types.

Consider smooth cubic curves in $\mathbb{P}^2(\mathbb{R})$. Each such cubic is either connected or has two connected components, and its Zariski closure in $\mathbb{P}^2(\mathbb{C})$ is also smooth. If C_1 and C_2 are smooth cubic curves in $\mathbb{P}^2(\mathbb{R})$, then $C_1 \times C_2$ can be oriented in such a way that for each regular map $\varphi\colon C_1 \times C_2 \to \mathbb{S}^2$, the topological degree $\deg(\varphi|_A)$ of the restriction of φ to a connected component A of $C_1 \times C_2$ does not depend on the choice of A. Moreover, the set

$$\mathrm{Deg}_{\mathcal{R}}(C_1, C_2) := \{m \in \mathbb{Z} \colon m = \deg(\psi|_A) \text{ for some regular map } \psi\colon C_1 \times C_2 \to \mathbb{S}^2\}$$

is a subgroup of $\mathbb{Z}$. These assertions are proved in Bochnak and Kucharz [1993, Theorem 3.1]. Define $b(C_1, C_2)$ to be the unique nonnegative integer satisfying

$$\mathrm{Deg}_{\mathcal{R}}(C_1, C_2) = b(C_1, C_2)\mathbb{Z}.$$

According to Hopf's theorem and Theorem 4.2, a continuous map $f\colon C_1 \times C_2 \to \mathbb{S}^2$ is homotopic to a regular map (or equivalently can be approximated by regular maps) if and only if for every connected component A of $C_1 \times C_2$, one has $\deg(f|_A) = b(C_1, C_2)r$ for some integer r independent of A. Thus, in this context, Problem 4.1 is reduced to the computation of the numerical invariant $b(C_1, C_2)$.

For any real number α in $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$, set

$$\tau_\alpha = \frac{1}{2}(1 + \alpha\sqrt{-1}) \text{ if } \alpha > 0 \text{ and } \tau_\alpha = \alpha\sqrt{-1} \text{ if } \alpha < 0.$$

The lattice $\Lambda_\alpha := \mathbb{Z} + \mathbb{Z}\tau_\alpha$ in $\mathbb{C}$ is stable under complex conjugation. Hence the numbers

$$g_2(\tau_\alpha) = 60 \sum_\omega \omega^{-4}, \quad g_3(\tau_\alpha) = 140 \sum_\omega \omega^{-6}$$

(summation over $\omega \in \Lambda_\alpha \setminus \{0\}$) are real and

$$D_\alpha := \{(x : y : z) \in \mathbb{P}^2(\mathbb{R}) \colon y^2 z = 4x^3 - g_2(\tau_\alpha)xz^2 - g_3(\tau_\alpha)z\}$$

is a smooth cubic curve in $\mathbb{P}^2(\mathbb{R})$. Each smooth cubic curve in $\mathbb{P}^2(\mathbb{R})$ is biregularly isomorphic to exactly one cubic D_α. Thus $\mathbb{R}^*$ can be regarded as a moduli space of smooth cubic curves in $\mathbb{P}^2(\mathbb{R})$. For $\alpha > 0$ (resp. $\alpha < 0$) the cubic D_α is connected (resp. has two connected components).

The invariant $b(D_{\alpha_1}, D_{\alpha_2})$ is explicitly computed in Bochnak and Kucharz [ibid.] for all pairs (α_1, α_2). In particular, it can take any nonnegative integer value. We recall only two cases.

First we deal with generic pairs (α_1, α_2).

Theorem 4.9. *For (α_1, α_2) in $\mathbb{R}^* \times \mathbb{R}^*$, the following conditions are equivalent:*

(a) *Each regular map from $D_{\alpha_1} \times D_{\alpha_2}$ into $\mathbb{S}^2$ is null homotopic.*

(b) $b(D_{\alpha_1}, D_{\alpha_2}) = 0$.

(c) *The product $\alpha_1\alpha_2$ is an irrational number.*

From the viewpoint of approximation, the following case is of greatest interest.

Theorem 4.10. *For (α_1, α_2) in $\mathbb{R}^* \times \mathbb{R}^*$, the following conditions are equivalent:*

(a) *Each continuous map from $D_{\alpha_1} \times D_{\alpha_2}$ into $\mathbb{S}^2$ can be approximated by regular maps.*

(b) $\alpha_1 > 0$, $\alpha_2 > 0$, *and* $b(D_{\alpha_1}, D_{\alpha_2}) = 1$.

(c) $\alpha_i = (p_i/q_i)\sqrt{d}$ *for* $i = 1, 2$, *where* p_i, q_i, d *are positive integers,* p_i *and* q_i *are relatively prime,* d *is square free,* $d \equiv 3 \pmod 4$, $p_1p_2q_1q_2 \equiv 1 \pmod 2$, *and* p_1p_2d *is divisible by* q_1q_2.

Theorems 4.9 and 4.10 show that a small perturbation of (α_1, α_2) can drastically change topological properties of regular maps from $D_{\alpha_1} \times D_{\alpha_2}$ into $\mathbb{S}^2$. Thus, in general, one cannot hope to find a comprehensive solution to Problem 4.1, even for X smooth with $\dim X = p$. It is therefore desirable to introduce maps which have good features of regular maps but are more flexible.

4.2 Approximation by regulous maps. Let X and $Y \subset \mathbb{R}^q$ be smooth real algebraic sets, and let k be a nonnegative integer. A map $f = (f_1, \dots, f_q) \colon X \to Y$ is said to be *k-regulous* if its components $f_i : X \to \mathbb{R}$ are k-regulous functions; 0-regulous maps are called *regulous*.

If f is a regulous map, we denote by $P(f)$ the smallest algebraic subset of X such that $f|_{X \setminus P(f)} \colon X \setminus P(f) \to Y$ is a regular map. Obviously, $P(f)$ is Zariski nowhere dense in X. We say that f is *nice* if $f(P(f)) \neq Y$.

We state the next result for $\mathcal{C}^\infty$ maps. This is convenient since such maps have regular values by Sard's theorem.

Theorem 4.11. *Let X be a compact smooth real algebraic set, $f : X \to \mathbb{S}^p$ a $\mathcal{C}^\infty$ map with $p \geq 1$, and $y \in \mathbb{S}^p$ a regular value of f. Assume that the $\mathcal{C}^\infty$ submanifold $f^{-1}(y)$ of X is isotopic to a smooth Zariski locally closed subset of X. Then:*

(i) *f is homotopic to a nice k-regulous map, where k is an arbitrary nonnegative integer.*

(ii) *f can be approximated by nice regulous maps.*

Theorem 4.11 is due to the first-named author. Part (i) is a simplified version of Kucharz [2009, Theorem 2.4]. The proof is based on the Pontryagin construction (framed cobordism), Łojasiewicz inequality, and Hironaka's resolution of singularities. In turn (ii) follows from Kucharz [2014a, Theorem 1.2] since, by Kucharz [1985, Theorem 2.1], $f^{-1}(y)$ can be approximated by smooth Zariski locally closed subsets of X.

Theorem 4.11 provides information also on continuous maps since they can be approximated by $\mathcal{C}^\infty$ maps.

According to Akbulut and King [1992, Theorem A], any compact $\mathcal{C}^\infty$ submanifold of $\mathbb{R}^n$ (resp. $\mathbb{S}^n$) is isotopic to a smooth Zariski locally closed subset of $\mathbb{R}^n$ (resp. $\mathbb{S}^n$). In particular, Theorem 4.11 yields the following.

Corollary 4.12. *Let (n, p) be a pair of positive integers. Then:*

(i) *Each continuous map from $\mathbb{S}^n$ into $\mathbb{S}^p$ is homotopic to a nice k-regulous map, where k is an arbitrary nonnegative integer.*

(ii) *Each continuous map from $\mathbb{S}^n$ into $\mathbb{S}^p$ can be approximated by nice regulous maps.*

With notation as in Theorem 4.11 we have $\dim f^{-1}(y) = \dim X - p$. Hence we get immediately the following.

Corollary 4.13. *Let X be a compact smooth real algebraic set of dimension p. Then:*

(i) *Each continuous map from X into $\mathbb{S}^p$ is homotopic to a nice k-regulous map, where k is an arbitrary nonnegative integer.*

(ii) *Each continuous map from X into $\mathbb{S}^p$ can be approximated by nice regulous maps.*

Comparing Theorems 4.8, 4.9 and 4.10 with Corollary 4.13 we see that k-regulous maps are indeed more flexible than regular ones. However, for each integer $p \geq 1$ there exist a compact smooth real algebraic set Y and a continuous map $g : Y \to \mathbb{S}^p$ such that $\dim Y = p + 1$ and g is not homotopic to any regulous map, cf. Kucharz and Kurdyka [2016b, Theorem 7.9]. In particular, Theorem 4.11 does not hold without some assumption on the $\mathcal{C}^\infty$ submanifold $f^{-1}(y) \subset X$. It would be very useful to formulate an appropriate assumption in terms of bordism. This is related to a certain conjecture, which has nothing to do with regulous maps and originates from the celebrated paper of Nash [1952] and the subsequent developments due to Tognoli [1973], Akbulut and King [1992], and other mathematicians.

For a real algebraic set X, a bordism class in the unoriented bordism group $\mathfrak{N}_*(X)$ is said to be *algebraic* if it can be represented by a regular map from a compact smooth real algebraic set into X.

Conjecture 4.14. *For any smooth real algebraic set X, the following holds: If M is a compact $\mathcal{C}^\infty$ submanifold of X and the unoriented bordism class of the inclusion map $M \hookrightarrow X$ is algebraic, then M is ε-isotopic to a smooth Zariski locally closed subset of X.*

Here "ε-isotopic" means isotopic via a $\mathcal{C}^\infty$ isotopy that can be chosen arbitrarily close, in the $\mathcal{C}^\infty$ topology, to the inclusion map. A slightly weaker assertion than the one in Conjecture 4.14 is known to be true: If the unoriented bordism class of the inclusion map $M \hookrightarrow X$ is algebraic, then the $\mathcal{C}^\infty$ submanifold $M \times \{0\}$ of $X \times \mathbb{R}$ is ε-isotopic to a smooth Zariski locally closed subset of $X \times \mathbb{R}$, cf. Akbulut and King [ibid., Theorem F].

Remark 4.15. Let X be a compact smooth real algebraic set and let $f : X \to \mathbb{S}^p$ be a continuous map. According to Kucharz and Kurdyka [2016a, Proposition 1.4], if Conjecture 4.14 holds, then the following conditions are equivalent:

(a) f is homotopic to a nice regulous map.

(b) f can be approximated by nice regulous maps.

Using a method independent of Conjecture 4.14, the first-named author proved in Kucharz [2016a] the following weaker result.

Theorem 4.16. *Let X be a compact smooth real algebraic set and let p be an integer satisfying* $\dim X + 3 \leq 2p$. *For a continuous map $f : X \to \mathbb{S}^p$, the following conditions are equivalent:*

(a) *f is homotopic to a nice regulous map.*

(b) *f can be approximated by nice regulous maps.*

Other results on topological properties of regulous maps can be found in Kucharz [2013b, 2014b, 2016a], Kucharz and Kurdyka [2016b], and Zieliński [2016].

5 Vector bundles

Let $\mathbb{F}$ stand for $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$ (the quaternions). We consider only left $\mathbb{F}$-vector spaces. When convenient, $\mathbb{F}$ will be identified with $\mathbb{R}^{d(\mathbb{F})}$, where $d(\mathbb{F}) = \dim_{\mathbb{R}} \mathbb{F}$.

5.1 Algebraic versus topological vector bundles. Let X be a real algebraic set. For any nonnegative integer n, let

$$\varepsilon_X^n(\mathbb{F}) = (X \times \mathbb{F}^n, p, X)$$

denote the product $\mathbb{F}$-vector bundle of rank n on X, where $X \times \mathbb{F}^n$ is regarded as a real algebraic set and $p\colon X \times \mathbb{F}^n \to X$ is the canonical projection.

An *algebraic* $\mathbb{F}$*-vector bundle on* X is an algebraic $\mathbb{F}$-vector subbundle of $\varepsilon_X^n(\mathbb{F})$ for some n (cf. Bochnak, Coste, and Roy [1998] for various characterizations of algebraic $\mathbb{F}$-vector bundles). The category of algebraic $\mathbb{F}$-vector bundles on X is equivalent to the category of finitely generated projective left $\mathcal{R}(X, \mathbb{F})$-modules, where $\mathcal{R}(X, \mathbb{F})$ is the ring of $\mathbb{F}$-valued regular functions on X.

Any algebraic $\mathbb{F}$-vector bundle on X can be regarded also as a topological $\mathbb{F}$-vector bundle. A topological $\mathbb{F}$-vector bundle is said to *admit an algebraic structure* if it is topologically isomorphic to an algebraic $\mathbb{F}$-vector bundle.

Problem 5.1. Which topological $\mathbb{F}$-vector bundles on X admit an algebraic structure?

It is convenient to bring into play the reduced Grothendieck group $\tilde{K}_{\mathbb{F}}(X)$ of topological $\mathbb{F}$-vector bundles on X. Since X has the homotopy type of a compact polyhedron Bochnak, Coste, and Roy [ibid.], the Abelian group $\tilde{K}_{\mathbb{F}}(X)$ is finitely generated Atiyah and Hirzebruch [1961] and Dyer [1969]. We let $\tilde{K}_{\mathbb{F}\text{-alg}}(X)$ denote the subgroup of $\tilde{K}_{\mathbb{F}}(X)$ generated by the classes of all topological $\mathbb{F}$-vector bundles on X that admit an algebraic structure.

If X is compact, then Problem 5.1 is equivalent to providing a description of $\tilde{K}_{\mathbb{F}\text{-alg}}(X)$. More precisely, the following holds.

Theorem 5.2. *Let X be a compact real algebraic set. Then:*

(i) *Two algebraic $\mathbb{F}$-vector bundles on X are algebraically isomorphic if and only if they are topologically isomorphic.*

(ii) *A topological $\mathbb{F}$-vector bundle on X admits an algebraic structure if and only if its class in $\tilde{K}_{\mathbb{F}}(X)$ belongs to $\tilde{K}_{\mathbb{F}\text{-alg}}(X)$.*

Theorem 5.2 follows from Swan [1977, Theorem 2.2], and a geometric proof is given in Bochnak, Coste, and Roy [1998]. Note that $\tilde{K}_{\mathbb{F}\text{-alg}}(X) = 0$ if and only if each algebraic $\mathbb{F}$-vector bundle on X is algebraically stably trivial. In turn, $\tilde{K}_{\mathbb{F}\text{-alg}}(X) = \tilde{K}_{\mathbb{F}}(X)$ if and only if each topological $\mathbb{F}$-vector bundle on X admits an algebraic structure.

According to Fossum [1969] and Swan [1977], we have the following.

Theorem 5.3. *For the unit n-sphere $\mathbb{S}^n$, the equality $\tilde{K}_{\mathbb{F}\text{-alg}}(\mathbb{S}^n) = \tilde{K}_{\mathbb{F}}(\mathbb{S}^n)$ holds.*

Benedetti and Tognoli [1980] proved that algebraization of topological vector bundles on a compact $\mathcal{C}^\infty$ manifold is always possible.

Theorem 5.4. *Let M be a compact $\mathcal{C}^\infty$ manifold. Then there exists an algebraic model X of M such that $\tilde{K}_{\mathbb{F}\text{-alg}}(X) = \tilde{K}_{\mathbb{F}}(X)$ for $\mathbb{F} = \mathbb{R}$, $\mathbb{F} = \mathbb{C}$ and $\mathbb{F} = \mathbb{H}$.*

The groups $\tilde{K}_{\mathbb{F}\text{-alg}}(-)$ have been extensively investigated by Bochnak and Kucharz [1989a, 1990, 1992] and Bochnak, Buchner, and Kucharz [1989]. In many cases, $\tilde{K}_{\mathbb{F}\text{-alg}}(-)$ are "small" subgroups of $\tilde{K}_{\mathbb{F}}(-)$. The following is a simplified version of Bochnak, Buchner, and Kucharz [ibid., Theorem 7.1].

Theorem 5.5. *Let M be a compact $\mathcal{C}^\infty$ submanifold of $\mathbb{R}^{n+1}$, with* $\dim M = n \geq 1$. *Then M is ε-isotopic to a smooth algebraic subset X of $\mathbb{R}^{n+1}$ such that the group $\tilde{K}_{\mathbb{F}\text{-alg}}(X)$ is finite for $\mathbb{F} = \mathbb{R}$, $\mathbb{F} = \mathbb{C}$ and $\mathbb{F} = \mathbb{H}$.*

The conclusion of Theorem 5.5 can be strengthened in some cases.

Example 5.6. Recall that $\tilde{K}_{\mathbb{F}}(\mathbb{S}^{4d}) = \mathbb{Z}$ for every positive integer d, cf. Husemoller [1975]. Hence, by Theorem 5.5, $\mathbb{S}^{4d}$ is ε-isotopic in $\mathbb{R}^{4d+1}$ to a smooth algebraic subset Σ^{4d} such that

$$\tilde{K}_{\mathbb{F}\text{-alg}}(\Sigma^{4d}) = 0 \text{ and } \tilde{K}_{\mathbb{F}}(\Sigma^{4d}) = \mathbb{Z}$$

for $\mathbb{F} = \mathbb{R}$, $\mathbb{F} = \mathbb{C}$ and $\mathbb{F} = \mathbb{H}$.

One readily sees that each topological $\mathbb{R}$-vector bundle on a smooth real algebraic curve admits an algebraic structure. Next we discuss some results on vector bundles on a product of real algebraic curves.

Example 5.7. Let $X = C_1 \times \cdots \times C_n$, where $C_1, \ldots, C_n$ are connected, compact smooth real algebraic curves. Then $\tilde{K}_{\mathbb{R}\text{-alg}}(X) = \tilde{K}_{\mathbb{R}}(X)$ if $n = 2$ or $n = 3$. This assertion is a special case of Bochnak and Kucharz [1989a, Theorem 1.6].

In Example 5.7, one cannot take $n \geq 4$.

Example 5.8. Let $\mathbb{T}^n := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ be the n-fold product of $\mathbb{S}^1$. According to Bochnak and Kucharz [1987b], $\tilde{K}_{\mathbb{C}\text{-alg}}(\mathbb{T}^n) = 0$. Furthermore, by Kucharz and Kurdyka [2016b, Example 1.11], we have $\tilde{K}_{\mathbb{R}\text{-alg}}(\mathbb{T}^n) \neq \tilde{K}_{\mathbb{R}}(\mathbb{T}^n)$ and $\tilde{K}_{\mathbb{H}\text{-alg}}(\mathbb{T}^n) \neq \tilde{K}_{\mathbb{H}}(\mathbb{T}^n)$ for $n \geq 4$.

Let $\mathbb{K}$ be a subfield of $\mathbb{F}$, where $\mathbb{K}$ (as $\mathbb{F}$) stands for $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. Any $\mathbb{F}$-vector bundle ξ can be regarded as a $\mathbb{K}$-vector bundle, which is indicated by $\xi_{\mathbb{K}}$.

Example 5.9. Let λ be a nontrivial topological $\mathbb{C}$-line bundle on $\mathbb{T}^2$. By Example 5.7, the $\mathbb{R}$-vector bundle $\lambda_{\mathbb{R}}$ on $\mathbb{T}^2$ admits an algebraic structure. However, in view of Example 5.8, λ does not admit an algebraic structure. The rth tensor power $\lambda^{\otimes r}$ is a nontrivial $\mathbb{C}$-line bundle, hence it does not admit an algebraic structure.

The next two theorems come from Bochnak and Kucharz [1992]. We use smooth real cubic curves $D_\alpha \subset \mathbb{P}^2(\mathbb{R})$, with $\alpha \in \mathbb{R}^*$, introduced in Section 4.1.

Theorem 5.10. *Let $X = D_\alpha \times \cdots \times D_\alpha$ be the n-fold product of D_α, where α is in $\mathbb{R}^*$ and $n \geq 2$. Then the following conditions are equivalent:*

(a) $\tilde{K}_{\mathbb{C}\text{-alg}}(X) = 0$.

(b) *The number α^2 is irrational.*

In this context, the equality $\tilde{K}_{\mathbb{C}\text{-alg}}(-) = \tilde{K}_{\mathbb{C}}(-)$ is characterized as follows.

Theorem 5.11. *Let $X = D_{\alpha_1} \times \cdots \times D_{\alpha_n}$, where $\alpha_1, \dots, \alpha_n$ are in $\mathbb{R}^*$ and $n \geq 2$. Then the following conditions are equivalent:*

(a) $\tilde{K}_{\mathbb{C}\text{-alg}}(X) = \tilde{K}_{\mathbb{C}}(X)$.

(b) $\alpha_i > 0$ *for all i, and* $b(D_{\alpha_i}, D_{\alpha_j}) = 1$ *for* $i \neq j$.

The pairs (α_i, α_j) with $b(D_{\alpha_i}, D_{\alpha_j}) = 1$ are explicitly described in Theorem 4.10.

In the next subsection we deal with vector bundles of a new type, which occupy an intermediate position between algebraic and topological vector bundles.

5.2 Regulous versus topological vector bundles. Let X be a real algebraic set. As in (3.8), by a stratification of X we mean a finite collection $\mathcal{S}$ of pairwise disjoint Zariski locally closed subsets whose union is X. An $\mathcal{S}$-*algebraic* $\mathbb{F}$-*vector bundle on* X is a topological $\mathbb{F}$-vector subbundle of $\varepsilon^n_X(\mathbb{F})$, for some n, such that the restriction $\xi|_S$ of ξ to each stratum $S \in \mathcal{S}$ is an algebraic $\mathbb{F}$-vector subbundle of $\varepsilon^n_S(\mathbb{F})$. If ξ and η are $\mathcal{S}$-algebraic $\mathbb{F}$-vector bundles on X, then an $\mathcal{S}$-*algebraic morphism* $\varphi\colon \xi \to \eta$ is a morphism of topological $\mathbb{F}$-vector bundles which induces a morphism of algebraic $\mathbb{F}$-vector bundles $\varphi_S\colon \xi|_S \to \eta|_S$ for each stratum $S \in \mathcal{S}$.

Definition 5.12. A *regulous* $\mathbb{F}$-*vector bundle on* X is an $\mathcal{S}$-algebraic $\mathbb{F}$-vector bundle for some stratification $\mathcal{S}$ of X. If ξ and η are regulous $\mathbb{F}$-vector bundles on X, then a *regulous morphism* $\varphi\colon \xi \to \eta$ is an $\mathcal{S}$-algebraic morphism for some stratification $\mathcal{S}$ of X such that both ξ and η are $\mathcal{S}$-algebraic $\mathbb{F}$-vector bundles.

In our joint paper Kucharz and Kurdyka [2016b], we introduced and investigated regulous (= stratified-algebraic) vector bundles. The main focus of Kucharz and Kurdyka [ibid.] and the subsequent papers Kucharz [2015b, 2016b], Kucharz and Kurdyka [2015], Kucharz and Zieliński [2018] is on comparison of algebraic, regulous, and topological vector bundles.

Regulous $\mathbb{F}$-vector bundles on X (together with regulous morphisms) form a category, which is equivalent to the category of finitely generated projective left $\mathcal{R}^0(X, \mathbb{F})$-modules, where $\mathcal{R}^0(X, \mathbb{F})$ is the ring of $\mathbb{F}$-valued regulous functions on X, cf. Kucharz and Kurdyka [2016b, Theorem 3.9].

A topological $\mathbb{F}$-vector bundle on X is said to *admit a regulous structure* if it is topologically isomorphic to a regulous $\mathbb{F}$-vector bundle. We let $\tilde{K}_{\mathbb{F}\text{-reg}}(X)$ denote the subgroup of $\tilde{K}_{\mathbb{F}}(X)$ generated by the classes of all topological $\mathbb{F}$-vector bundles on X that admit a regulous structure.

We have the following counterpart of Theorem 5.2, cf. Kucharz and Kurdyka [ibid.].

Theorem 5.13. *Let X be a compact real algebraic set. Then:*

(i) *Two regulous $\mathbb{F}$-vector bundles on X are regulously isomorphic if and only if they are topologically isomorphic.*

(ii) *A topological $\mathbb{F}$-vector bundle on X admits a regulous structure if and only if its class in $\tilde{K}_{\mathbb{F}}(X)$ belongs to $\tilde{K}_{\mathbb{F}\text{-reg}}(X)$.*

Hence $\tilde{K}_{\mathbb{F}\text{-reg}}(X) = \tilde{K}_{\mathbb{F}}(X)$ if and only if each topological $\mathbb{F}$-vector bundle on X admits a regulous structure.

The following result of Kucharz and Kurdyka [ibid.] should be compared with Theorem 5.3 and Example 5.6.

Theorem 5.14. *Let X be a compact real algebraic set that is homotopically equivalent to the unit n-sphere $\mathbb{S}^n$. Then $\tilde{K}_{\mathbb{F}\text{-reg}}(X) = \tilde{K}_{\mathbb{F}}(X)$.*

In contrast to Example 5.8 and Theorems 5.10 and 5.11, we proved in Kucharz and Kurdyka [ibid.] the following.

Theorem 5.15. *Let $X = X_1 \times \cdots \times X_n$, where X_i is a compact real algebraic set that is homotopically equivalent to the unit d_i-sphere $\mathbb{S}^{d_i}$ for $1 \leq i \leq n$. Then*

$$2\tilde{K}_{\mathbb{R}}(X) \subset \tilde{K}_{\mathbb{R}\text{-reg}}(X),\ \tilde{K}_{\mathbb{C}\text{-reg}}(X) = \tilde{K}_{\mathbb{C}}(X) \text{ and } \tilde{K}_{\mathbb{H}\text{-reg}}(X) = \tilde{K}_{\mathbb{H}}(X).$$

It is possible that $\tilde{K}_{\mathbb{R}\text{-reg}}(X) = \tilde{K}_{\mathbb{R}}(X)$ in Theorem 5.15, but no proof is available even for $X = \mathbb{T}^n$ with $n \geq 4$.

Our next result comes from Kucharz and Kurdyka [2015].

Theorem 5.16. *Let X be a compact real algebraic set that is homotopically equivalent to $\mathbb{S}^{d_1} \times \cdots \times \mathbb{S}^{d_n}$. Then the quotient group $\tilde{K}_{\mathbb{F}}(X)/\tilde{K}_{\mathbb{F}\text{-reg}}(X)$ is finite.*

If $n \geq 5$, then there exists an algebraic model X of $\mathbb{T}^n$ with $\tilde{K}_{\mathbb{F}}(X)/\tilde{K}_{\mathbb{F}\text{-reg}}(X) \neq 0$ for $\mathbb{F} = \mathbb{R}$, $\mathbb{F} = \mathbb{C}$ and $\mathbb{F} = \mathbb{H}$, cf. Kucharz and Kurdyka [2016b, Example 7.10].

For any real algebraic set X, we let $\tilde{K}_{\mathbb{F}}^{(\mathrm{crk})}(X)$ denote the subgroup of $\tilde{K}_{\mathbb{F}}(X)$ generated by the classes of all topological $\mathbb{F}$-vector bundles of constant rank. Define $\Gamma_{\mathbb{F}}(X)$ to be the quotient group

$$\Gamma_{\mathbb{F}}(X) := \tilde{K}_{\mathbb{F}}^{(\mathrm{crk})}(X)/(\tilde{K}_{\mathbb{F}\text{-reg}}(X) \cap \tilde{K}_{\mathbb{F}}^{(\mathrm{crk})}(X))$$

(cf. Kucharz and Kurdyka [2015] for an equivalent description). Evidently, $\Gamma_{\mathbb{F}}(X) = \tilde{K}_{\mathbb{F}}(X)/\tilde{K}_{\mathbb{F}\text{-reg}}(X)$ if X is connected. Note, however, that for the real algebraic curve C of (1.5), one has $\Gamma_{\mathbb{R}}(C) = 0$, while the group $\tilde{K}_{\mathbb{R}}(C)/\tilde{K}_{\mathbb{R}\text{-reg}}(C)$ is infinite.

Conjecture 5.17. *For any compact real algebraic set X, the group $\Gamma_{\mathbb{F}}(X)$ is finite.*

In Kucharz and Kurdyka [ibid.], we proved that Conjecture 5.17 holds in low dimensions.

Theorem 5.18. *If X is a compact real algebraic set of dimension at most 8, then the group $\Gamma_{\mathbb{F}}(X)$ is finite.*

For $\mathbb{C}$-line bundles, the following is expected.

Conjecture 5.19. *For any compact real algebraic set X and any topological $\mathbb{C}$-line bundle λ on X, the $\mathbb{C}$-line bundle $\lambda^{\otimes 2}$ admits a regulous structure.*

According to Kucharz [2015b], Conjecture 4.14 implies Conjecture 5.19. If $\dim X \leq 8$, then for some positive integer r, the $\mathbb{C}$-line bundle $\lambda^{\otimes r}$ admits a regulous structure, cf. Kucharz and Kurdyka [2015]. This should be compared with Example 5.9.

The key role in the proofs of the results presented in this subsection plays the following theorem of Kucharz and Kurdyka [2016b].

Theorem 5.20. *Let X be a compact real algebraic set. A topological $\mathbb{F}$-vector bundle ξ on X admits a regulous structure if and only if the $\mathbb{R}$-vector bundle $\xi_{\mathbb{R}}$ admits a regulous structure.*

By Example 5.9, one cannot substitute "algebraic" for "regulous" in Theorem 5.20.

References

Janusz Adamus and Hadi Seyedinejad (2017). "A proof of Kurdyka's conjecture on arc-analytic functions". *Math. Ann.* 369.1-2, pp. 387–395. MR: 3694650 (cit. on pp. 739, 743).

Selman Akbulut and Henry King (1992). "On approximating submanifolds by algebraic sets and a solution to the Nash conjecture". *Invent. Math.* 107.1, pp. 87–98. MR: 1135465 (cit. on pp. 756, 757).

Michael F. Atiyah and Friedrich Hirzebruch (1961). "Vector bundles and homogeneous spaces". In: *Proc. Sympos. Pure Math., Vol. III*. American Mathematical Society, Providence, R.I., pp. 7–38. MR: 0139181 (cit. on p. 758).

James Ax (1969). "Injective endomorphisms of varieties and schemes". *Pacific J. Math.* 31, pp. 1–7. MR: 0251036 (cit. on p. 742).

Riccardo Benedetti and Alberto Tognoli (1980). "On real algebraic vector bundles". *Bull. Sci. Math. (2)* 104.1, pp. 89–112. MR: 560747 (cit. on p. 759).

Andrzej Białynicki-Birula and Maxwell Rosenlicht (1962). "Injective morphisms of real algebraic varieties". *Proc. Amer. Math. Soc.* 13, pp. 200–203. MR: 0140516 (cit. on p. 742).

Edward Bierstone and Pierre D. Milman (1990). "Arc-analytic functions". *Invent. Math.* 101.2, pp. 411–424. MR: 1062969 (cit. on pp. 739, 744).

– (1997). "Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant". *Invent. Math.* 128.2, pp. 207–302. MR: 1440306 (cit. on p. 744).

Edward Bierstone, Pierre D. Milman, and Adam Parusiński (1991). "A function which is arc-analytic but not continuous". *Proc. Amer. Math. Soc.* 113.2, pp. 419–423. MR: 1072083 (cit. on pp. 739, 743).

Marcin Bilski, Wojciech Kucharz, Anna Valette, and Guillaume Valette (2013). "Vector bundles and regulous maps". *Math. Z.* 275.1-2, pp. 403–418. MR: 3101814 (cit. on p. 737).

Zbigniew Błocki (1992). "Singular sets of separately analytic functions". *Ann. Polon. Math.* 56.2, pp. 219–225. MR: 1159993 (cit. on p. 745).

Jacek Bochnak, Michael Buchner, and Wojciech Kucharz (1989). "Vector bundles over real algebraic varieties". *K-Theory* 3.3, pp. 271–298. MR: 1040403 (cit. on p. 759).

Jacek Bochnak, Michel Coste, and Marie-Françoise Roy (1998). *Real Algebraic Geometry*. Vol. 36. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin, pp. x+430. MR: 1659509 (cit. on pp. 738, 740, 747, 750, 752, 758).

Jacek Bochnak and Wojciech Kucharz (1987a). "Algebraic approximation of mappings into spheres". *Michigan Math. J.* 34.1, pp. 119–125. MR: 873026 (cit. on pp. 751, 752).

– (1987b). "Realization of homotopy classes by algebraic mappings". *J. Reine Angew. Math.* 377, pp. 159–169. MR: 887406 (cit. on pp. 751–753, 759).

– (1988). "On real algebraic morphisms into even-dimensional spheres". *Ann. of Math. (2)* 128.2, pp. 415–433. MR: 960952 (cit. on pp. 751–753).

– (1989a). "*K*-theory of real algebraic surfaces and threefolds". *Math. Proc. Cambridge Philos. Soc.* 106.3, pp. 471–480. MR: 1010372 (cit. on p. 759).

Jacek Bochnak and Wojciech Kucharz (1989b). "Algebraic models of smooth manifolds". *Invent. Math.* 97.3, pp. 585–611. MR: 1005007 (cit. on pp. 751–753).

– (1990). "On vector bundles and real algebraic morphisms". In: *Real analytic and algebraic geometry (Trento, 1988)*. Vol. 1420. Lecture Notes in Math. Springer, Berlin, pp. 65–71. MR: 1051204 (cit. on pp. 753, 759).

– (1991a). "Nonisomorphic algebraic models of a smooth manifold". *Math. Ann.* 290.1, pp. 1–2. MR: 1107659 (cit. on p. 753).

– (1991b). "Polynomial mappings from products of algebraic sets into spheres". *J. Reine Angew. Math.* 417, pp. 135–139. MR: 1103909 (cit. on pp. 751, 752).

– (1992). "Vector bundles on a product of real cubic curves". *K-Theory* 6.6, pp. 487–497. MR: 1204824 (cit. on pp. 759, 760).

– (1993). "Elliptic curves and real algebraic morphisms". *J. Algebraic Geom.* 2.4, pp. 635–666. MR: 1227471 (cit. on pp. 751–755).

– (1999). "The Weierstrass approximation theorem for maps between real algebraic varieties". *Math. Ann.* 314.4, pp. 601–612. MR: 1709103 (cit. on pp. 751, 752).

– (2007). "Real algebraic morphisms represent few homotopy classes". *Math. Ann.* 337.4, pp. 909–921. MR: 2285743 (cit. on pp. 751, 752).

Jacek Bochnak, Wojciech Kucharz, and Robert Silhol (1997). "Morphisms, line bundles and moduli spaces in real algebraic geometry". *Inst. Hautes Études Sci. Publ. Math.* 86, 5–65 (1998). MR: 1608561 (cit. on p. 752).

Salomon Bochner and William T. Martin (1948). *Several Complex Variables*. Princeton Mathematical Series, vol. 10. Princeton University Press, Princeton, N. J., pp. ix+216. MR: 0027863 (cit. on p. 745).

Armand Borel (1969). "Injective endomorphisms of algebraic varieties". *Arch. Math. (Basel)* 20, pp. 531–537. MR: 0255545 (cit. on p. 742).

Eldon Dyer (1969). *Cohomology theories*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., New York-Amsterdam, pp. xiii+183. MR: 0268883 (cit. on p. 758).

Charles Fefferman and János Kollár (2013). "Continuous solutions of linear equations". In: *From Fourier analysis and number theory to Radon transforms and geometry*. Vol. 28. Dev. Math. Springer, New York, pp. 233–282. MR: 2986959 (cit. on p. 748).

Goulwen Fichou (2005). "Motivic invariants of arc-symmetric sets and blow-Nash equivalence". *Compos. Math.* 141.3, pp. 655–688. MR: 2135282 (cit. on pp. 739, 744).

Goulwen Fichou, Johannes Huisman, Frédéric Mangolte, and Jean-Philippe Monnier (2016). "Fonctions régulues". *J. Reine Angew. Math.* 718, pp. 103–151. MR: 3545880 (cit. on pp. 737, 746, 747, 750).

Goulwen Fichou, Jean-Philippe Monnier, and Ronan Quarez (2017). "Continuous functions on the plane regular after one blowing-up". *Math. Z.* 285.1-2, pp. 287–323. MR: 3598813 (cit. on p. 737).

Robert M. Fossum (1969). "Vector bundles over spheres are algebraic". *Invent. Math.* 8, pp. 222–225. MR: 0250298 (cit. on p. 758).

Toshizumi Fukui, Satoshi Koike, and Tzee-Char Kuo (1998). "Blow-analytic equisingularities, properties, problems and progress". In: *Real analytic and algebraic singularities (Nagoya/Sapporo/Hachioji, 1996)*. Vol. 381. Pitman Res. Notes Math. Ser. Longman, Harlow, pp. 8–29. MR: 1607662 (cit. on p. 744).

Riccardo Ghiloni (2006). "On the space of morphisms into generic real algebraic varieties". *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 5.3, pp. 419–438. MR: 2274786 (cit. on p. 752).

– (2007). "Second order homological obstructions on real algebraic manifolds". *Topology Appl.* 154.17, pp. 3090–3094. MR: 2355513 (cit. on p. 752).

Heisuke Hironaka (1964). "Resolution of singularities of an algebraic variety over a field of characteristic zero. I". *Ann. of Math. (2)* 79, pp. 109–203. MR: 0199184 (cit. on p. 747).

Dale Husemoller (1975). *Fibre bundles*. Second Edition. Graduate Texts in Mathematics, No. 20. Springer-Verlag, New York-Heidelberg, pp. xv+327. MR: 0370578 (cit. on p. 759).

Nuria Joglar-Prieto and Frédéric Mangolte (2004). "Real algebraic morphisms and del Pezzo surfaces of degree 2". *J. Algebraic Geom.* 13.2, pp. 269–285. MR: 2047699 (cit. on p. 752).

Satoshi Koike and Adam Parusiński (2003). "Motivic-type invariants of blow-analytic equivalence". *Ann. Inst. Fourier (Grenoble)* 53.7, pp. 2061–2104. MR: 2044168 (cit. on pp. 739, 744).

János Kollár, Wojciech Kucharz, and Krzysztof Kurdyka (2018). "Curve-rational functions". *Math. Ann.* 370.1-2, pp. 39–69. MR: 3747483 (cit. on pp. 737, 744, 746, 748, 749).

János Kollár and Krzysztof Nowak (2015). "Continuous rational functions on real and p-adic varieties". *Math. Z.* 279.1-2, pp. 85–97. MR: 3299844 (cit. on pp. 738, 746–748).

Wojciech Kucharz (1985). "On homology of real algebraic sets". *Invent. Math.* 82.1, pp. 19–25. MR: 808106 (cit. on p. 756).

– (2005). "Homology classes represented by semialgebraic arc-symmetric sets". *Bull. London Math. Soc.* 37.4, pp. 514–524. MR: 2143731 (cit. on pp. 740, 742).

– (2009). "Rational maps in real algebraic geometry". *Adv. Geom.* 9.4, pp. 517–539. MR: 2574137 (cit. on pp. 737, 746, 747, 750, 756).

– (2010). "Complex cycles on algebraic models of smooth manifolds". *Math. Ann.* 346.4, pp. 829–856. MR: 2587094 (cit. on pp. 752, 753).

– (2013a). "Regular maps into real Fermat varieties". *Bull. Lond. Math. Soc.* 45.5, pp. 1086–1090. MR: 3105000 (cit. on p. 752).

Wojciech Kucharz (2013b). "Regular versus continuous rational maps". *Topology Appl.* 160.12, pp. 1375–1378. MR: 3072698 (cit. on pp. 738, 746, 757).
– (2014a). "Approximation by continuous rational maps into spheres". *J. Eur. Math. Soc. (JEMS)* 16.8, pp. 1555–1569. MR: 3262451 (cit. on pp. 738, 746, 756).
– (2014b). "Continuous rational maps into the unit 2-sphere". *Arch. Math. (Basel)* 102.3, pp. 257–261. MR: 3181715 (cit. on pp. 738, 746, 757).
– (2015a). "Complex cycles as obstructions on real algebraic varieties". *Glasg. Math. J.* 57.2, pp. 343–347. MR: 3333945 (cit. on p. 752).
– (2015b). "Some conjectures on stratified-algebraic vector bundles". *J. Singul.* 12, pp. 92–104. MR: 3317142 (cit. on pp. 738, 746, 760, 762).
– (2016a). "Continuous rational maps into spheres". *Math. Z.* 283.3-4, pp. 1201–1215. MR: 3520001 (cit. on pp. 738, 746, 757).
– (2016b). "Stratified-algebraic vector bundles of small rank". *Arch. Math. (Basel)* 107.3, pp. 239–249. MR: 3538519 (cit. on pp. 738, 760).
Wojciech Kucharz and Krzysztof Kurdyka (2015). "Comparison of stratified-algebraic and topological K-theory". arXiv: 1511.04238 (cit. on pp. 738, 746, 760–762).
– (2016a). "Some conjectures on continuous rational maps into spheres". *Topology Appl.* 208, pp. 17–29. MR: 3506967 (cit. on pp. 738, 746, 757).
– (2016b). "Stratified-algebraic vector bundles". *J. Reine Angew. Math.* (Cit. on pp. 738, 746, 756, 757, 759–762).
– (2017). "Linear equations on real algebraic surfaces". *Manuscripta Math.* 154.3-4, pp. 285–296. MR: 3713914 (cit. on pp. 738, 746, 748).
Wojciech Kucharz and Maciej Zieliński (2018). "Regulous vector bundles". *Math. Nachr.* (Cit. on pp. 738, 760).
Tzee Char Kuo (1985). "On classification of real singularities". *Invent. Math.* 82.2, pp. 257–262. MR: 809714 (cit. on p. 744).
Krzysztof Kurdyka (1988). "Ensembles semi-algébriques symétriques par arcs". *Math. Ann.* 282.3, pp. 445–462. MR: 967023 (cit. on pp. 739–741, 743).
– (1991). "A counterexample to subanalyticity of an arc-analytic function". In: *Proceedings of the Tenth Conference on Analytic Functions (Szczyrk, 1990)*. Vol. 55, pp. 241–243. MR: 1141439 (cit. on pp. 740, 743).
– (1994). "An arc-analytic function with nondiscrete singular set". *Ann. Polon. Math.* 59.3, pp. 251–254. MR: 1282776 (cit. on pp. 740, 743).
– (1999). "Injective endomorphisms of real algebraic sets are surjective". *Math. Ann.* 313.1, pp. 69–82. MR: 1666793 (cit. on pp. 740, 742).
Krzysztof Kurdyka and Adam Parusiński (2007). "Arc-symmetric sets and arc-analytic mappings". In: *Arc spaces and additive invariants in real algebraic and analytic geometry*. Vol. 24. Panor. Synthèses. Soc. Math. France, Paris, pp. 33–67. MR: 2409688 (cit. on pp. 740, 742).

– (2012). "On the non-analyticity locus of an arc-analytic function". *J. Algebraic Geom.* 21.1, pp. 61–75. MR: 2846679 (cit. on pp. 740, 744).

Krzysztof Kurdyka and Kamil Rusek (1988). "Surjectivity of certain injective semialgebraic transformations of $\mathbf{R}^n$". *Math. Z.* 200.1, pp. 141–148. MR: 972400 (cit. on pp. 740, 742).

Jean-Louis Loday (1973). "Applications algébriques du tore dans la sphère et de $S^p \times S^q$ dans S^{p+q}", 79–91. Lecture Notes in Mathematics, Vol. 342. MR: 0368034 (cit. on p. 752).

Clint McCrory and Adam Parusiński (2003). "Virtual Betti numbers of real algebraic varieties". *C. R. Math. Acad. Sci. Paris* 336.9, pp. 763–768. MR: 1989277 (cit. on pp. 740, 744).

Jean-Philippe Monnier (2018). "Semi-algebraic geometry with rational continuous functions". *Math. Ann.* (Cit. on p. 738).

John Nash (1952). "Real algebraic manifolds". *Ann. of Math. (2)* 56, pp. 405–421. MR: 0050928 (cit. on pp. 741, 753, 757).

Donald J. Newman (1960). "One-one polynomial maps". *Proc. Amer. Math. Soc.* 11, pp. 867–870. MR: 0122817 (cit. on p. 742).

Yildiray Ozan (1995). "On entire rational maps in real algebraic geometry". *Michigan Math. J.* 42.1, pp. 141–145. MR: 1322195 (cit. on p. 752).

Adam Parusiński (1994). "Subanalytic functions". *Trans. Amer. Math. Soc.* 344.2, pp. 583–595. MR: 1160156 (cit. on pp. 740, 744).

– (2001). "On the preparation theorem for subanalytic functions". In: *New developments in singularity theory*. Vol. 21. NATO Science Series II: Mathematics, Physics and Chemistry. Kluwer Academic Publishers, Dordrecht, pp. viii+472. MR: 1849301 (cit. on pp. 740, 744).

– (2004). "Topology of injective endomorphisms of real algebraic sets". *Math. Ann.* 328.1-2, pp. 353–372. MR: 2030380 (cit. on pp. 740, 742).

Adam Parusiński and Laurenţiu Păunescu (2017). "Arc-wise analytic stratification, Whitney fibering conjecture and Zariski equisingularity". *Adv. Math.* 309, pp. 254–305. MR: 3607278 (cit. on pp. 740, 744).

Jiagui Peng and Zizhou Tang (1999). "Algebraic maps from spheres to spheres". *Sci. China Ser. A* 42.11, pp. 1147–1154. MR: 1749862 (cit. on p. 752).

Józef Siciak (1990). "Singular sets of separately analytic functions". *Colloq. Math.* 60/61.1, pp. 281–290. MR: 1096378 (cit. on p. 745).

Richard G. Swan (1977). "Topological examples of projective modules". *Trans. Amer. Math. Soc.* 230, pp. 201–234. MR: 0448350 (cit. on p. 758).

René Thom (1954). "Quelques propriétés globales des variétés différentiables". *Comment. Math. Helv.* 28, pp. 17–86. MR: 0061823 (cit. on p. 742).

Alberto Tognoli (1973). "Su una congettura di Nash". *Ann. Scuola Norm. Sup. Pisa (3)* 27, pp. 167–185. MR: 0396571 (cit. on pp. 753, 757).

Francisco-Javier Turiel (2007). "Polynomial maps and even dimensional spheres". *Proc. Amer. Math. Soc.* 135.8, pp. 2665–2667. MR: 2302590 (cit. on p. 752).

Reginald Wood (1968). "Polynomial maps from spheres to spheres". *Invent. Math.* 5, pp. 163–168. MR: 0227999 (cit. on p. 752).

Maciej Zieliński (2016). "Homotopy properties of some real algebraic maps". *Homology Homotopy Appl.* 18.1, pp. 287–294. MR: 3498647 (cit. on pp. 738, 746, 757).

Received 2017-12-05.

Wojciech Kucharz
Institute of Mathematics
Faculty of Mathematics and Computer Science
Jagiellonian University
Łojasiewicza 6
30-348 Kraków
Poland
Wojciech.Kucharz@im.uj.edu.pl

Krzysztof Kurdyka
Univ. Grenoble Alpes,
Univ. Savoie Mont Blanc, CNRS,
LAMA, 73000 Chambéry
France
Krzysztof.Kurdyka@univ-savoie.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (767–798)

PAIRS OF INVARIANTS OF SURFACE SINGULARITIES

András Némethi

Abstract

We discuss several invariants of complex normal surface singularities with a special emphasis on the comparison of analytic–topological pairs of invariants. Additionally we also list several open problems related with them.

1 Introduction

Singularity theory aims to study the singular points of algebraic/analytic varieties. It was born together with the classical algebraic geometry, but step by step became an independent discipline within algebraic and complex geometry. Furthermore, it also created formidable connections with other fields, like topology or differential equations. By crucial classification projects in mainstream mathematics (e.g. the Mori program in algebraic geometry targeting classification of varieties, or developments in low-dimensional topology) singularities theory became even more a central area.

In both local and global case the study of surfaces became central. In the global case, after the 'classical' classification of Enriques and the Italian school, a 'modern' classification was provided by Kodaira in 60's based on sheaves, cohomologies and characteristic classes with special emphasis on the relationships of the analytic structures with invariants of the underlying smooth 4–manifolds: e.g. Hodge or Riemann–Roch–Hirzebruch formulas. Crucial open questions were formulated targeting this type of ties: e.g. the purely topological characterization of rational surfaces (as an addendum of Castelnuovo's criterion) or of K3 surfaces (asked by Kodaira). The appearance of Donaldson and Seiberg–Witten theories gave a powerful impetus of such comparison results, and generated a series of new open questions. These were inherited by local singularity theory too. The lack of

The author was partially supported by NKFIH Grant 112735.
MSC2010: primary 14B05; secondary 14J17, 32S25, 57M27, 57R57, 32S45.
Keywords: normal surface singularities, rational singularities, elliptic singularities, $\mathbb{Q}$–homology spheres, geometric genus, link of singularities, divisorial filtration, Poincaré series, Seiberg–Witten invariants of 3–manifolds, Casson Invariant Conjecture, Seiberg–Witten Invariant Conjecture, Heegaard–Floer homology, graded roots, surgery 3-manifolds, unicuspidal rational projective plane curves.

classification results in 3 and 4 dimensional topology put even more in the highlight the possible analytic/algebraic connections.

This is the guiding principle of the present manuscript too: what are the ties between analytic and topological structures of a local complex normal surface singularity.

It is always exciting to understand such bridges between topology and rigid analytic/algebraic geometry. In this note by the style of the presentation we even try to emphasize more the existing parallel pairs of invariants, theorems and constructions from both sides. Though the presented list definitely is not exhaustive, it supports rather well the philosophical conviction that any result in the analytic part must have a topological counterpart, and vice versa.

The guiding pair is $\mathcal{P}(\mathbf{t}) \leftrightarrow Z(\mathbf{t})$, where $\mathcal{P}(\mathbf{t})$ is the multivariable Poincaré series associated with the divisorial filtration given by the irreducible curves of a resolution, while $Z(\mathbf{t})$ is a combinatorial 'zeta' function read from the corresponding resolution graph.

2 Normal surface singularities. Analytic Invariants

2.1 Definitions, notations. Let (X,o) be a complex normal surface singularity. Let $\pi : \widetilde{X} \to X$ be a good resolution with dual graph Γ whose vertices are denoted by $\mathcal{V}$. Set $E := \pi^{-1}(o)$. Let M be the link of (X,o), and we will assume that M is a rational homology sphere. This happens if and only if Γ is a tree and all the irreducible exceptional curves $\{E_v\}_{v\in\mathcal{V}}$ have genus 0.

Set $L := H_2(\widetilde{X},\mathbb{Z})$. It is freely generated by the classes of the irreducible exceptional curves. If L' denotes $H^2(\widetilde{X},\mathbb{Z})$, then the intersection form $(\,,\,)$ on L provides an embedding $L \hookrightarrow L'$ with factor the first homology group H of the link. (In fact, L' is the dual lattice of $(L,(\,,\,))$.) Moreover, $(\,,\,)$ extends to L'. L' is freely generated by the duals E_v^*, where $(E_v^*, E_w) = -1$ for $v = w$ and $= 0$ else.

Effective classes $l = \sum r_v E_v \in L'$ with all $r_v \in \mathbb{Q}_{\geq 0}$ are denoted by $L'_{\geq 0}$ and $L_{\geq 0} := L'_{\geq 0} \cap L$. Denote by $\mathcal{S}'$ the (Lipman's) anti-nef cone $\{l' \in L' \;:\; (l', E_v) \leq 0 \text{ for all } v\}$. It is generated over $\mathbb{Z}_{\geq 0}$ by the base-elements E_v^*. Since all the entries of E_v^* are strict positive, $\mathcal{S}'$ is a sub-cone of $L'_{\geq 0}$, and for any fixed $a \in L'$ the set $\{l' \in \mathcal{S}' \;:\; l' \ngeq a\}$ is finite. Set $\mathfrak{C} := \{\sum l'_v E_v \in L',\ 0 \leq l'_v < 1\}$. For any $l' \in L'$ write its class in H by $[l']$, and or any $h \in H$ let $r_h \in L'$ be its unique representative in $\mathfrak{C}$. Denote by $\theta : H \to \widehat{H}$ the isomorphism $[l'] \mapsto e^{2\pi i (l',\cdot)}$ of H with its Pontrjagin dual $\widehat{H}$.

Denote by $K \in L'$ the *canonical class* satisfying $(K+E_v, E_v) = -2$ for all $v \in \mathcal{V}$. We set $\chi(l') = -(l', l'+K)/2$; by Riemann-Roch theorem $\chi(l) = \chi(\mathcal{O}_l)$ for any $l \in L_{>0}$.

Most of the analytic geometry of $\widetilde{X}$ is described by its line bundles and their cohomology groups. E.g., the geometric genus of (X,o) is $p_g := h^1(\widetilde{X}, \mathcal{O}_{\widetilde{X}})$. In this note we mostly target the following numerical invariants (below $\mathcal{L} \in \mathrm{Pic}(\widetilde{X})$ and $l \in L_{>0}$):

$$(2.1.1)\qquad (a)\ \dim\ H^0(\mathcal{L})\,/\,H^0(\mathcal{L}(-l)) \quad \text{and} \quad (b)\ \dim H^1(\mathcal{L}).$$

Their behaviour for arbitrary line bundles is rather complicated, however for *natural line bundles* we have several (sometimes even topological) descriptions/characterizations. These line bundles are provided by the splitting of the cohomological exponential exact sequence Némethi [n.d.(b), §3]:

$$0 \to H^1(\widetilde{X}, \mathcal{O}_{\widetilde{X}}) \to \mathrm{Pic}(\widetilde{X}) \xrightarrow{c_1} L' \to 0.$$

The first Chern class c_1 has an obvious section on the subgroup L, namely $l \mapsto \mathcal{O}_{\widetilde{X}}(l)$. This section has a unique extension $\mathcal{O}(\cdot)$ to L'. We call a line bundle *natural* if it is in the image of this section.

One can recover these bundles via coverings as follows. Let $c : (Y, o) \to (X, o)$ be the universal abelian covering of (X, o), $\pi_Y : \widetilde{Y} \to Y$ the normalized pullback of π by c, and $\widetilde{c} : \widetilde{Y} \to \widetilde{X}$ the morphism which covers c. Then the action of H on (Y, o) lifts to $\widetilde{Y}$ and one has an H–eigensheaf decomposition (Némethi [ibid., (3.7)] or Okuma [2008, (3.5)]):

$$(2.1.2)\qquad \widetilde{c}_*\mathcal{O}_{\widetilde{Y}} = \bigoplus_{l' \in \mathfrak{C}} \mathcal{O}(-l') \quad (\mathcal{O}(-l') \text{ being the } \theta([l'])\text{-eigenspace of } \widetilde{c}_*\mathcal{O}_{\widetilde{Y}}).$$

Note that the geometric genus of Y is $h^1(\widetilde{Y}, \mathcal{O}_{\widetilde{Y}})$, hence, by Equation (2.1.2), $\{h^1(\widetilde{X}, \mathcal{O}_{\widetilde{X}}(-r_h))\}_{h\in H}$ are the dimensions of the H–eigenspaces; we call them equivariant geometric genera of (X, o).

2.2 Series associated with the divisorial filtration. For those natural line bundles which appear in Equation (2.1.2), the dimensions from Equation (2.1.1)(a) can be organized in a generating function. Indeed, once a resolution π is fixed, $\mathcal{O}_{Y,o}$ inherits the *divisorial multi-filtration* (cf. Némethi [2008b, (4.1.1)]):

$$(2.2.1)\qquad \mathcal{F}(l') := \{f \in \mathcal{O}_{Y,o} \mid \mathrm{div}(f \circ \pi_Y) \geq \widetilde{c}^*(l')\}.$$

Let $\mathfrak{h}(l')$ be the dimension of the $\theta([l'])$-eigenspace of $\mathcal{O}_{Y,o}/\mathcal{F}(l')$. Then, one defines the *equivariant divisorial Hilbert series* by

$$(2.2.2)\qquad \mathcal{H}(\mathbf{t}) = \sum_{l' \in L'} \mathfrak{h}(l') t_1^{l'_1} \cdots t_s^{l'_s} = \sum_{l' \in L'} \mathfrak{h}(l')\mathbf{t}^{l'} \in \mathbb{Z}[[L']] \ \ (l' = \sum_i l'_i E_i).$$

Notice that the terms of the sum reflect the H-eigenspace decomposition too: $\mathfrak{h}(l')\mathbf{t}^{l'}$ contributes to the $\theta([l'])$-eigenspace. For example, $\sum_{l\in L} \mathfrak{h}(l)\mathbf{t}^l$ corresponds to the H-invariants, hence it is the *Hilbert series* of $\mathcal{O}_{X,o}$ associated with the $\pi^{-1}(o)$-divisorial multi-filtration (considered and intensively studied, see e.g. Cutkosky, Herzog, and Reguera [2004] and the citations therein, or Campillo, Delgado, and Gusein-Zade [2004]).

The 'graded version' associated with the Hilbert series is defined (cf. Campillo, Delgado, and Gusein-Zade [2004] and Guseĭn-Zade, Delgado, and Kampil'o [2008]) as

$$\mathcal{P}(\mathbf{t}) = -\mathcal{H}(\mathbf{t}) \cdot \prod_v (1 - t_v^{-1}) \in \mathbb{Z}[[L']]. \tag{2.2.3}$$

Although the multiplication by $\prod_v (1 - t_v^{-1})$ in $\mathbb{Z}[[L']]$ is not injective, hence apparently $\mathcal{P}$ contains less information then $\mathcal{H}$, they, in fact, determine each other as we will see in Equation (2.4.2).

If we write the series $\mathcal{P}(\mathbf{t})$ as $\sum_{l'} \mathfrak{p}(l')\mathbf{t}^{l'}$, then

$$\mathfrak{p}(l') = \sum_{I \subseteq \mathcal{V}} (-1)^{|I|+1} \dim \frac{H^0(\widetilde{X}, \mathcal{O}(-l'))}{H^0(\widetilde{X}, \mathcal{O}(-l' - E_I))} \tag{2.2.4}$$

and $\mathcal{P}$ is supported in the cone $\mathcal{S}'$.

2.3 Quasipolynomials and the periodic constants associated with series. The following definitions are motivated by properties of Hilbert–Samuel functions and also by Ehrhart theory and the properties of its qusipolynomials. The periodic constant of one–variable series was introduced in Némethi and Okuma [2009], Okuma [2008], and Braun and Némethi [2010], the multivariable generalization is treated in László and Némethi [2014].

Let $S(t) = \sum_{l \geq 0} c_l t^l \in \mathbb{Z}[[t]]$ be a formal power series with one variable. Assume that for some $p \in \mathbb{Z}_{>0}$ the counting function $Q^{(p)}(n) := \sum_{l=0}^{pn-1} c_l$ is a polynomial $\mathfrak{Q}^{(p)}$ in n. Then the constant term $\mathfrak{Q}^{(p)}(0)$ is independent of p and it is called the *periodic constant* $\mathrm{pc}(S)$ of the series S. E.g., if $S(t)$ is a finite polynomial, then $\mathrm{pc}(S)$ exists and it equals $S(1)$. If the coefficients of $S(t)$ are given by a Hilbert function $l \mapsto c(l)$, which admits a Hilbert polynomial $H(l)$ with $c(l) = H(l)$ for $l \gg 0$, then $S^{reg}(t) = \sum_{l \geq 0} H(l)t^l$ has zero periodic constant and $\mathrm{pc}(S) = \mathrm{pc}(S - S^{reg}) + \mathrm{pc}(S^{reg}) = (S - S^{reg})(1)$, measuring the difference between the Hilbert function and Hilbert polynomial.

For the multivariable case we consider a (negative) definite lattice $L = \mathbb{Z}\langle E_v \rangle_v$, its dual lattice L', and a series $S(\mathbf{t}) \in \mathbb{Z}[[L']]$, $S(\mathbf{t}) = \sum_{l' \in L'} s(l')\mathbf{t}^{l'}$. We decompose S as $S = \sum_{h \in H} S_h$, where $S_h(\mathbf{t}) = \sum_{[l']=h} s(l')\mathbf{t}^{l'}$, and we consider the following 'counting function of the coefficients'

$$Q_h : L'_h := \{x \in L' : [x] = h\} \to \mathbb{Z}, \quad Q_h(x) = \sum_{l' \not\geq x,\, [l']=h} s(l'). \tag{2.3.1}$$

Assume that there exist a real cone $\mathcal{K} \subset L' \otimes \mathbb{R}$ whose affine closure is top–dimensional, $l'_* \in \mathcal{K}$, a sublattice $\widetilde{L} \subset L$ of finite index, and a quasipolynomial $\mathfrak{Q}_h(l)$ $(l \in \widetilde{L})$ such that $Q_h(l + r_h) = \mathfrak{Q}_h(l)$ for any $l + r_h \in (l'_* + \mathcal{K}) \cap (\widetilde{L} + r_h)$. Then we say that the counting function Q_h (or just $S_h(\mathbf{t})$) admits a quasipolynomial in $\mathcal{K}$, namely $\mathfrak{Q}_h(l)$, and also an (equivariant, multivariable) *periodic constant* associated with $\mathcal{K}$, which is defined by

$$\mathrm{pc}^{\mathcal{K}}(S_h(\mathbf{t})) := \mathfrak{Q}_h(0). \tag{2.3.2}$$

The definition does not depend on the choice of the sublattice $\widetilde{L}$, which corresponds to the choice of p in the one–variable case. This is responsible for the name 'periodic' in the definition. The definition is independent of the choice of l'_* as well.

By general theory of multivariable Ehrhart-type quasipolynomials, for a nicely defined series one can construct a conical chamber decomposition of the space $L' \otimes \mathbb{R}$, such that each cone satisfies the above definition (hence provides a periodic constant), for details see László and Némethi [ibid.] or Szenes and Vergne [2003]. However, it turns out, that in all our situations the whole $\mathcal{S}'_{\mathbb{R}}$ (the real Lipman cone) will be a unique chamber.

2.4 The quasipolynomial of $\mathcal{P}$. Let $\mathcal{P}$ and $\mathcal{H}$ be the series defined in SubSection 2.2.

Theorem 2.4.1. *For any $l' \in L'$ one has*

$$\mathfrak{h}(l') = \sum_{a \in L,\, a \ngeq 0} \mathfrak{p}(l' + a). \tag{2.4.2}$$

Furthermore, there exists a constant $\mathrm{const}_{[-l']}$, *depending only on the class of* $[-l'] \in H$, *such that*

$$-h^1(\widetilde{X}, \mathcal{O}(-l')) = \sum_{a \in L,\, a \ngeq 0} \mathfrak{p}(l' + a) + \mathrm{const}_{[-l']} + \frac{(K + 2l')^2 + |\mathcal{V}|}{8}. \tag{2.4.3}$$

By taking $l' = r_h$ one can identify $\mathrm{const}_{[-l']}$ with the equivariant geometric genus, that is $-h^1(\widetilde{X}, \mathcal{O}(-r_h)) = \mathrm{const}_{-h} + ((K + 2r_h)^2 + |\mathcal{V}|)/8$. In particular, the coefficients of the series $\mathcal{P}(\mathbf{t})$ determine the invariants Equation (2.1.1)((a)-(b)) for all natural line bundles.

Write $l' = l + r_h$ for $l \in L$. Note that if $l \in L_{\leq 0}$, then in Equation (2.4.3) the sum will not appear (check the support of $\mathcal{P}$). On the other hand, if in Equation (2.4.3) $l' \in -K + \mathcal{S}'$ then by the vanishing of $h^1(\mathcal{O}(-l'))$ we get that $\sum_{a \in L,\, a \ngeq 0} \mathfrak{p}(l' + a)$ is the multivariable quadratic function $\chi(l) - (r_h, l) + h^1(\widetilde{X}, \mathcal{O}(-r_h))$. This quadratic function is the quasipolynomial of $\mathcal{P}$ in $\mathcal{K} = \mathcal{S}'_{\mathbb{R}}$, and its periodic constant is $h^1(\widetilde{X}, \mathcal{O}(-r_h))$.

2.5 Reductions of $\mathcal{P}$. Let $\mathcal{U}$ be a nonempty subset of $\mathcal{V}$, and fix $h \in H$ as well. One can introduce the reduction of the series $\mathcal{P}_h$ to the variables $\{t_v\}_{v\in\mathcal{U}}$ in two different ways. First, we can consider the multivariable divisorial filtration induced by the exceptional divisors $\{E_v\}_{v\in\mathcal{U}}$, define the Hilbert series $\mathcal{H}_h^{\mathcal{U}}$ and $\mathcal{P}_h^{\mathcal{U}}$ in variables $\{t_v\}_{v\in\mathcal{U}}$ and associated with $\theta(h)$–eigendecomposition similarly as $\mathcal{H}_h$ and $\mathcal{P}_h$ was defined in Section 2.2. The second possibility is to reduced the variables in the original $\mathcal{P}_h$. The point is that the two constructions have the same output: $\mathcal{P}_h^{\mathcal{U}} = \mathcal{P}_h(\mathbf{t})|_{t_v=1 \text{ for all } v\notin\mathcal{U}}$. In this article we mostly discuss the case $\mathcal{U} = \{v\}$ for a certain vertex v, however all the discussions can be extended for arbitrary $\mathcal{U}$.

2.6 Surgery formulae for $h^1(\widetilde{X}, \mathcal{L})$. Let (X,o) and $\pi : \widetilde{X} \to X$ as above. We fix a vertex $v \in \mathcal{V}$. Let $\cup_{j\in J}\Gamma_j$ be the connected components of the graph obtained from Γ by deleting v and its adjacent edges. Let X' be the space obtained from $\widetilde{X}$ by contracting all irreducible exceptional curves except E_v to normal points. It has $|J|$ normal singular points $\{o_j\}_j$, the images of the connected components of $E - E_v$. Let X_j be a small Stein neighbourhood of o_j in X', $\widetilde{X}_j = \tau^{-1}(X_j)$ its pre-image via the contraction $\tau : \widetilde{X} \to X'$, and $\tau(E) = E' \subset X'$. We denote the local singularities by (X_j, o_j).

We say that the Assumption (*C*) is satisfied if $nE' \subset X'$ is a Cartier divisor for certain $n > 0$.

Theorem 2.6.1. *Okuma [2008] Set $\mathcal{U} = \{v\}$ and fix $h \in H$. Under the Assumption (C)*

$$h^1(\widetilde{X}, \mathcal{O}_{\widetilde{X}}(-r_h)) = \mathrm{pc}(\mathcal{P}_h^{\{v\}}(t_v)) + \sum_j h^1(\widetilde{X}_j, \mathcal{O}_{\widetilde{X}}(-r_h)|_{\widetilde{X}_j}).$$

E.g., this applied for $\mathcal{O}_{\widetilde{X}}$ gives $p_g(X,o) = \mathrm{pc}(\mathcal{P}_h^{\{v\}}(t_v)) + \sum_j p_g(X_j, o_j)$. In general, $\mathcal{O}_{\widetilde{X}}(-r_h)|_{\widetilde{X}_j}$ is not a natural line bundle on $\widetilde{X}_j$, however e.g. for splice quotient singularities it is the line bundle associated with the cohomology restriction of $-r_h$ (for details see Section 3.3). Hence, Theorem 2.6.1 provides an ideal inductive procedure for the computation of the cohomology of natural line bundles.

3 Topological invariants. The series $Z(\mathbf{t})$

3.1 The series $Z(\mathbf{t})$. The *multivariable topological series* is the Taylor expansion $Z(\mathbf{t}) = \sum_{l'} z(l')\mathbf{t}^{l'} \in \mathbb{Z}[[L']]$ at the origin of the rational function

$$f(\mathbf{t}) = \prod_{v\in\mathcal{V}} (1 - \mathbf{t}^{E_v^*})^{\delta_v - 2}. \tag{3.1.1}$$

It is supported in $\mathcal{S}'$. Similarly as $\mathcal{P}$, it decomposes as $Z(\mathbf{t}) = \sum_{h\in H} Z_h(\mathbf{t})$. In fact,

$$Z_h(\mathbf{t}) := \frac{1}{|H|} \cdot \sum_{\rho\in\widehat{H}} \rho(h)^{-1} \cdot \prod_{v\in\mathcal{V}} \left(1-\rho([E_v^*])\mathbf{t}^{E_v^*}\right)^{\delta_v-2}. \tag{3.1.2}$$

3.2 Seiberg–Witten invariants of the link M. Let $\widetilde{\sigma}_{can}$ be the *canonical $spin^c$–structure on $\widetilde{X}$* identified by $c_1(\widetilde{\sigma}_{can}) = -K$, and let $\sigma_{can} \in \mathrm{Spin}^c(M)$ be its restriction to M, called the *canonical $spin^c$–structure on M*. $\mathrm{Spin}^c(M)$ is an H–torsor with action denoted by $*$.

We denote by $\mathfrak{sw}_\sigma(M) \in \mathbb{Q}$ the *Seiberg–Witten invariant* of M indexed by the $spin^c$–structures $\sigma \in \mathrm{Spin}^c(M)$ (cf. Lim [2000] and Nicolaescu [2004]). We will use the sign convention of Braun and Némethi [2010] and Némethi [2011].

In the last years several combinatorial expressions were established for the Seiberg–Witten invariants. For rational homology spheres, Nicolaescu [2004] showed that $\mathfrak{sw}(M)$ is equal to the Reidemeister–Turaev torsion normalized by the Casson–Walker invariant. In the case when M is a negative definite plumbed rational homology sphere, combinatorial formula for Casson–Walker invariant in terms of the plumbing graph can be found in Lescop [1996], and the Reidemeister–Turaev torsion is determined by Némethi and Nicolaescu [2002] using Dedekind–Fourier sums. A different combinatorial formula of $\{\mathfrak{sw}_\sigma(M)\}_\sigma$ was proved in Némethi [2011] using qualitative properties of the coefficients of the series $Z(\mathbf{t})$.

Theorem 3.2.1. *Némethi [ibid.] The counting function of $Z_h(\mathbf{t})$ in the cone $S'_{\mathbb{R}}$ admits the (quasi)polynomial*

$$\mathfrak{Q}_h(l) = -\frac{(K+2r_h+2l)^2+|\mathcal{V}|}{8} - \mathfrak{sw}_{-h*\sigma_{can}}(M), \tag{3.2.2}$$

whose periodic constant is

$$\mathrm{pc}^{S'_{\mathbb{R}}}(Z_h(\mathbf{t})) = \mathfrak{Q}_h(0) = -\mathfrak{sw}_{-h*\sigma_{can}}(M) - \frac{(K+2r_h)^2+|\mathcal{V}|}{8}. \tag{3.2.3}$$

The right hand side of Equation (3.2.3) with opposite sign is called the *r_h–normalized Seiberg–Witten invariant* of M.

3.3 Surgery formulae for the normalized Seiberg–Witten invariants. Surgery formulae for a certain 3–manifold invariant, in general, compare the invariant of M with the invariants of different surgery modifications of M. In the case of plumbed 3–manifolds, one compares the invariants associated with 3–manifolds obtained by different modifications of the graph. The ‘standard’ topological surgery formulae for the Seiberg–Witten invariant (induced by exact triangles of certain cohomology theories, cf. Ozsváth and Szabó [2004a], Greene [2013], and Némethi [2011]) compare the invariants of three such 3–manifolds (see also Section 5.1 here). Furthermore, in these approaches, one cannot separate a certain fixed $spin^c$ structure, the theory mixes always several of them. (See also Turaev [2001].) The next formula is different: it compares the Seiberg–Witten invariant of two 3–manifolds via the periodic constant of a series, and they split according to the $spin^c$–structures.

Let us fix $v \in \mathcal{V}$ and consider the notations of Section 2.6. For each $j \in J$ we consider the inclusion operator $\iota_j : L(\Gamma_j) \to L(\Gamma)$, $E_v(\Gamma_j) \mapsto E_v(\Gamma)$; let $\iota_j^* : L'(\Gamma) \to L'(\Gamma_j)$ be its dual (the cohomological restriction defined by $\iota_j^*(E_v^*(\Gamma)) = E_v^*(\Gamma_j)$ if $v \in \mathcal{V}(\Gamma_j)$, and $= 0$ otherwise).

Next, consider an arbitrary $spin^c$–structure $\widetilde{\sigma}$ on $\widetilde{X}$. Since $\mathrm{Spin}^c(\widetilde{X})$ is an L'–torsor, there is a unique $l' \in L'$ such that $\widetilde{\sigma} = l' * \widetilde{\sigma}_{can}$. Its restriction to $\mathrm{Spin}^c(M)$ is $\sigma = [l'] * \sigma_{can}$. We write $\widetilde{\sigma}_j$ the restriction of $\widetilde{\sigma}$ to each $\widetilde{X}_j$. Since the canonical $spin^c$–structure of $\widetilde{X}$ restricts to the canonical $spin^c$–structure $\widetilde{\sigma}_{can,j}$ of $\widetilde{X}_j$, $\widetilde{\sigma} = l' * \widetilde{\sigma}_{can}$ restricts to $\widetilde{\sigma}_j := \iota_j^*(l') * \widetilde{\sigma}_{can,j} \in \mathrm{Spin}^c(\widetilde{X}_j)$, whose restriction to the boundary $M_j = M(X_j, o_j) = \partial \widetilde{X}_j$ is $\sigma_j = [\iota_j^*(l')] * \sigma_{can,j}$.

Theorem 3.3.1. *Braun and Némethi [2010] Fix $\mathcal{U} = \{v\}$ and $h \in H$. Extend $h * \sigma_{can} \in \mathrm{Spin}^c(M)$ as $\widetilde{\sigma} := r_h * \widetilde{\sigma}_{can} \in \mathrm{Spin}^c(\widetilde{X})$ and consider the corresponding restrictions. Then*

$$\mathfrak{sw}_{-h*\sigma_{can}}(M) + \frac{(K+2r_h)^2 + |\mathcal{V}|}{8} =$$
$$= \sum_j \Big(\mathfrak{sw}_{-[\iota_j^*(r_h)]*\sigma_{can,j}}(M_j) + \frac{(K(\Gamma_j) + 2\iota_j^*(r_h))^2 + |\mathcal{V}(\Gamma_j))|}{8} \Big) - \mathrm{pc}(Z_h^{\{v\}}(t_v)).$$

For a generalization to an arbitrary $\mathcal{U}$ and to an arbitrary extension $\widetilde{\sigma} := l' * \widetilde{\sigma}_{can}$ see László, Nagy, and Némethi [n.d.].

4 Topological invariants. The lattice cohomology of M

The $spin^c$–structures of M can also be indexed as follows. Set $\mathrm{Char} = \{k \in L' : (k + x, x) \in 2\mathbb{Z}$ for all $x \in L\}$, the set of characteristic elements of $\widetilde{X}$. Then L acts on Char by $l * k = k + 2l$, and the set of orbits $[k] = k + 2L$ is an H torsor identified with $\mathrm{Spin}^c(M)$.

For any $k \in \mathrm{Char}$ one also defines $\chi_k : L' \to \mathbb{Q}$ by $\chi_k(l') := -(l', l' + k)/2$. We write χ for χ_K.

The Seiberg–Witten invariant is the (normalized) Euler-characteristic of the Seiberg–Witten monopole Floer homology of Kronheimer–Mrowka, or equivalently, of the Heegaard–Floer homology of Ozsváth and Szabó. These theories had an extreme influence on the modern mathematics, solving (or disproving) a long list of old conjectures (e.g. Thom Conjecture, or conjectures regarding classification of 4-manifolds, or famous old problems in knot theory); see the long list of distinguished articles of Kronheimer–Mrowka or Ozsváth–Szabó. In Ozsváth and Szabó [2003b] Ozsváth and Szabó provided a computation of the Heegaard–Floer homology for some special plumbed 3-manifolds. This computation resonated incredibly with the theory of computation sequences used in Artin–Laufer program (see e.g. Laufer [1977] and Némethi [1999a,b]). These two facts influenced considerably the definition of the lattice cohomology.

4.1 Short review of Heegaard–Floer homology $HF^+(M)$. We assume that M is an oriented rational homology 3–sphere, and we restrict ourselves to the $+$–theory of Ozsváth and Szabó. The Heegaard–Floer homology $HF^+(M)$ is a $\mathbb{Z}[U]$–module with a $\mathbb{Q}$–grading compatible with the $\mathbb{Z}[U]$–action, where $\deg(U)=-2$. Additionally, $HF^+(M)$ has another $\mathbb{Z}_2$–grading; $HF^+(M)_{even}$, respectively $HF^+(M)_{odd}$ denote the graded parts. Moreover, $HF^+(M)$ has a natural direct sum decomposition of $\mathbb{Z}[U]$–modules (compatible with all the gradings): $HF^+(M)=\oplus_\sigma HF^+(M,\sigma)$ indexed by the $spin_c$–structures σ of M. For any σ one has

$$HF^+(M,\sigma)=\mathcal{T}^+_{d(M,\sigma)}\oplus HF^+_{red}(M,\sigma),$$

a graded $\mathbb{Z}[U]$–module isomorphism, where $\mathcal{T}^+_r$ denotes $\mathbb{Z}[U^{-1}]$ as a $\mathbb{Z}[U]$–module, in which the degree of 1 is r; and $HF^+_{red}(M,\sigma)$ has finite $\mathbb{Z}$–rank and an induced $\mathbb{Z}_2$–grading. One also considers

$$\chi(HF^+(M,\sigma)) := \operatorname{rank}_{\mathbb{Z}} HF^+_{red,even}(M,\sigma)-\operatorname{rank}_{\mathbb{Z}} HF^+_{red,odd}(M,\sigma).$$

Then via $\chi(HF^+(M,\sigma))-d(M,\sigma)/2$ one gets the Seiberg–Witten invariant of (M,σ). By changing the orientation one has $\chi(HF^+(M,\sigma))=-\chi(HF^+(-M,\sigma))$ and $d(M,\sigma)=-d(-M,\sigma)$.

4.2 Lattice cohomology of M. Now we review some facts from the lattice cohomology theory, introduced by the author in Némethi [2008a]. The construction captures the structure of lattice points inside of some real ellipsoids. $L\otimes\mathbb{R}$ has a natural cellular decomposition into cubes. The set of zero–dimensional cubes is provided by the lattice points L. Any $l\in L$ and subset $I\subset\mathcal{V}$ of cardinality q defines a q–dimensional cube (l,I), which has its vertices in the lattice points $(l+\sum_{i\in I'}E_j)_{I'}$, where I' runs over all subsets of I.

Next, we fix $k \in \mathrm{Char}$, we set $\chi_k(l) = -(l, l+k)/2$ and $m_k := \min_{l \in L} \chi_k(l)$. Finally, for any fixed integer $n \geq m_k$ we denote by S_n the union of all q–cubes in the real ellipsoid $\{l \in L \otimes \mathbb{R} \,:\, \chi_k(l) \leq n\}$. Then one defines

$$\mathbb{H}^p(\Gamma, k) := \oplus_{n \geq m_k} H^p(S_n, \mathbb{Z}).$$

For each fixed p, the module $\mathbb{H}^p$ is in a natural way $\mathbb{Z}$– (in fact, $2\mathbb{Z}$)–graded: $H^p(S_n, \mathbb{Z})$ consists of the $2n$–homogeneous elements. Also, it is a $\mathbb{Z}[U]$–module; the U–action is induced by the restriction $H^p(S_{n+1}, \mathbb{Z}) \to H^p(S_n, \mathbb{Z})$. Moreover, there is an augmentation decomposition

$$\mathbb{H}^0(\Gamma, k) = (\oplus_{n \geq m_k} \mathbb{Z}) \oplus (\oplus_{n \geq m_k} \widetilde{H}^0(S_n, \mathbb{Z})) = \mathcal{T}^+_{2m_k} \oplus \mathbb{H}^0_{red}(\Gamma, k).$$

This module $\mathbb{H}^*(\Gamma, k)$ is independent of the choice of the resolution (or plumbing) graph Γ (hence depends only on the 3-manifold M), and it depends only on the class $[k] = k + 2L$ (i.e. only on the corresponding $spin^c$–structure) up to a shift in grading. In order to fix one module in each class we take one $k' \in [k]$ with $m_{k'} = 0$, and we set $\mathbb{H}^*(\Gamma, [k]) := \mathbb{H}(\Gamma, k')$ (which is independent of the choice). For any rational number r we denoted by $\mathbb{H}^*(\Gamma, [k])[r]$ the module isomorphic to $\mathbb{H}^*(\Gamma, [k])$, but whose grading is shifted by r. (The $(d+r)$–homogeneous part of $\mathbb{H}^*(\Gamma, [k])[r]$ is isomorphic with the d–homogeneous part of $\mathbb{H}^*(\Gamma, [k])$.) It is also convenient to redefine $\mathbb{H}^p_{red} := \mathbb{H}^p$ for $p \geq 1$; this is motivated by the fact that $\mathbb{H}^*_{red} = \oplus_{p \geq 0} \mathbb{H}^p_{red}$ has finite $\mathbb{Z}$–rank.

5 The relations of the lattice cohomology with other invariants

5.1 Exact sequence. Let us fix a vertex v_0, we consider the graphs $\Gamma \setminus v_0$ and $\Gamma^+_{v_0}$, where the first one is obtained from Γ by deleting the vertex v_0 and adjacent edges, while the second one is obtained from Γ by increasing the decoration of the vertex v_0 by 1. We will assume that $\Gamma^+_{v_0}$ is still negative definite. Then there exists an exact sequence of $\mathbb{Z}[U]$–modules of the following type:

$$\cdots \longrightarrow \mathbb{H}^q(\Gamma^+_{v_0}) \longrightarrow \mathbb{H}^q(\Gamma) \longrightarrow \mathbb{H}^q(\Gamma \setminus v_0) \longrightarrow \mathbb{H}^{q+1}(\Gamma^+_{v_0}) \longrightarrow \cdots$$

The first 3 terms of the exact sequence (i.e. the $\mathbb{H}^0$–part) appeared in Ozsváth and Szabó [2003b] and Némethi [2005], the exact sequence over $\mathbb{Z}_2$–coefficients was proved in Greene [2013], the general case in Némethi [2011]. For more properties see Némethi [ibid.].

5.2 Rational singularities. By definition, (X, o) is rational if $p_g(X, o) = 0$. This is an analytic property, however, Artin replaced the vanishing of p_g by a topological criterion formulated in terms of χ: (X, o) is rational if and only if $\chi(l) > 0$ for any $l \in L_{>0}$ (independently of the resolution) Artin [1962, 1966]. For a different, the so–called 'Laufer's rationality criterion', see Laufer [1972]. Any connected negative definite graph with this property is called rational graph. The set of rational graphs include e.g. all the graphs with the following property: if e_v and δ_v denote the Euler decoration and the valency of a vertex, then one requires $-e_v \geq \delta_v$ for any $v \in \mathcal{V}$. Hence, if we decrease sufficiently the Euler decorations of any graph we get a rational graph. The class of rational graphs is closed while taking subgraphs and decreasing the Euler numbers.

Rationality in terms of the lattice cohomology is characterized as follows:

Theorem 5.2.1. *Némethi [2005] The following facts are equivalent:*

(a) Γ *is rational;*

(b) $\mathbb{H}^0(\Gamma, K) = \mathcal{T}_0^+$*;*

(c) $\mathbb{H}^0(\Gamma, K) = \mathcal{T}_m^+$ *for some* $m \in \mathbb{Z}$*; or equivalently,* $\mathbb{H}^0_{red}(\Gamma, K) = 0$*;*

(d) $\mathbb{H}^*_{red}(\Gamma, k) = 0$ *for all* $k \in \text{Char}$.

Moreover, if Γ *is rational then* $\min \chi_{k_r} = 0$.

5.3 (Weakly) elliptic singularities. A normal surface singularity (X, o) is called *elliptic*, if (one of its) resolution graph is elliptic. A graph Γ is elliptic if $\min_{l>0} \chi(l) = 0$ (cf. Wagreich [1970] and Laufer [1977], see also Némethi [1999b]).

The set of elliptic singularities includes all the singularities with $p_g = 1$, and all the Gorenstein singularities with $p_g = 2$. But an elliptic singularity might have arbitrary large p_g.

Ellipticity in terms of the lattice cohomology is characterized as follows:

Theorem 5.3.1. *The following facts are equivalent:*

(a) Γ *is elliptic;*

(b) $\mathbb{H}^0(\Gamma, K) = \mathcal{T}_0^+ \oplus \big(\mathcal{T}_0(1)\big)^{\oplus \ell}$ *for some* $\ell \geq 1$*. Here* $\mathcal{T}_0(1)$ *is a free* $\mathbb{Z}$*–module of rank one with trivial U–action and concentrated at degree 0. The integer* ℓ *above can be identified with the length of the elliptic sequence. In particular, if the graph* Γ *is minimally elliptic then* $\ell = 1$.

5.4 'Bad' vertices, AR graphs. We fix an integer $n \geq 0$. We say that a negative definite graph has at most n 'bad' vertices if we can find n vertices $\{v_k\}_{1 \leq k \leq n}$, such that by decreasing their Euler decorations we get a rational graph (this fact makes sense because of Section 5.2). In general, the choice of the bad vertices is not unique. A graph with at most one bad vertex is called *almost rational*, or AR, cf. Némethi [2005, 2008a]. Here are some AR graphs:

1) All rational and elliptic graphs are AR.

2) Any star-shaped graph is AR (modify the central vertex).

3) The rational surgery 3–manifolds $S^3_{-p/q}(\mathcal{K})$ ($\mathcal{K}$ algebraic knot, $p/q > 0$) are AR.

4) The class of AR graphs is closed while taking subgraphs and decreasing the Euler numbers.

Theorem 5.4.1. *Némethi [2011] If Γ has at most n bad vertices then $\mathbb{H}^q_{red}(\Gamma) = 0$ for $q \geq n$. In particular, if Γ is AR then $\mathbb{H}^0(\Gamma)$ is the only nonzero module.*

If Γ has at most $n \geq 2$ bad vertices $\{v_k\}_{1\leq k\leq n}$ such that $\Gamma \setminus v_1$ has at most $(n-2)$ bad vertices then $\mathbb{H}^q_{red}(\Gamma) = 0$ for $q \geq n-1$.

See Némethi [2005, (8.2)(5.b)] for a graph Γ with 2 bad vertices $\{v_1, v_2\}$ such that $\Gamma \setminus v_1$ has only rational components.

5.5 Relation with Heegaard–Floer theory. In Némethi [2008a] the author formulated the following

Conjecture 5.5.1. *For any plumbed rational homology sphere associated with a connected negative definite graph Γ, and for any $k \in$ Char, one has*

$$d(M,[k]) = \max_{k'\in[k]} \frac{(k')^2 + |\mathcal{V}|}{4} = \frac{k^2 + |\mathcal{V}|}{4} - 2\cdot \min \chi_k .$$

Furthermore,

$$HF^+_{red,even}(-M,[k]) = \bigoplus_{p\ even} \mathbb{H}^p_{red}(\Gamma,[k])[-d], \ and$$
$$HF^+_{red,odd}(-M,[k]) = \bigoplus_{p\ odd} \mathbb{H}^p_{red}(\Gamma,[k])[-d].$$

Both parts of the Conjecture were verified for almost rational graphs in Némethi [ibid.], for two bad vertices in Ozsváth, Stipsicz, and Szabó [2014], see Némethi [2008a, p. 8.4] too. Otherwise, the Conjecture is still open.

Note that (conjecturally) $\mathbb{H}^*$ has a richer structure: its p–filtration $\mathbb{H}^* = \oplus_p \mathbb{H}^p$ collapses at the level of HF^+ to a $\mathbb{Z}_2$ odd/even filtration.

5.6 Relation with Seiberg–Witten invariant. For any $k \in$ Char the (normalized) Euler characteristic of the lattice cohomology is defined as

$$\mathfrak{eu}(\mathbb{H}^*(\Gamma,k)) := -\min \chi_k + \sum_p (-1)^p \operatorname{rank} \mathbb{H}^p_{red}(\Gamma,k).$$

Then, it turns out that the (normalized) Euler characteristic of the lattice cohomology equals the (normalized) Seiberg–Witten invariant Némethi [2011] (this fact supports Conjecture 5.5.1 too):

$$-\mathfrak{sw}_{[k]}(M) - \frac{k^2 + |\mathcal{V}|}{8} = -\min \chi_k + \sum_p (-1)^p \operatorname{rank} \mathbb{H}^p_{red}(\Gamma, k).$$

5.7 Relation with L–spaces. By Section 5.2 Γ is rational if and only if $\mathbb{H}^*_{red}(\Gamma) = 0$. On the other hand, following Ozsváth and Szabó, M is an L–space by definition if and only if $HF^+_{red} = 0$. Their equivalence is predicted by Conjecture 5.5.1; in fact this 'tip of the iceberg' statement was proved in Némethi [2017]:

Theorem 5.7.1. *The following facts are equivalent:*
(i) (X, o) is a rational singularity (or, Γ is a rational graph),
(ii) the link M is an L–space.

$(i) \Rightarrow (ii)$ follows from lattice cohomology theory Némethi [2005, 2008a], while $(ii) \Rightarrow (i)$ uses partly the following equivalence $(ii) \Leftrightarrow (iii)$, where (iii) means that $\pi_1(M)$ is not a left–orderable group. The equivalence $(ii) \Leftrightarrow (iii)$ was proved in Hanselman, J. Rasmussen, S. D. Rasmussen, and Watson [n.d.] for any graph–manifolds. For arbitrary 3–manifolds was conjectured by Boyer, Gordon, and Watson [2013], for different developments and other references see Hanselman, J. Rasmussen, S. D. Rasmussen, and Watson [n.d.] and Némethi [2017].

Problem 5.7.2. Characterize elliptic singularities by a certain property of $\pi_1(M)$.

5.8 Reductions. In the lattice cohomology computations it is convenient to take a special representative k for the $spin^c$–structure $[k]$. Indeed, if $s_h \in L'$ is the minimal representative of $h \in H$ in $\mathcal{S}'$, and we take $k_r := K + 2s_h$ as representative for $[k]$, and we define the weighted cubes with the weight function χ_{k_r}, then one has the following 'homotopical identity': $H^p(S_n, \mathbb{Z}) = H^p(S_n \cap L_{\geq 0}, \mathbb{Z})$, where $S_n \cap L_{\geq 0}$ denotes the subset of S_n consisting of cubes with all vertices in $L_{\geq 0}$. Hence, with the natural notations (after notational modification $\mathbb{H}^*(L, k) = \mathbb{H}^*(\Gamma, k)$) we have

Theorem 5.8.1. *László and Némethi [2015]* $\mathbb{H}^*(L, \chi_{k_r}) = \mathbb{H}^*(L_{\geq 0}, \chi_{k_r})$.

This can be reduced even further. Fix k_r as above, that is $k_r = K + 2s_h$ for some h, and rewrite s_h as $s_{[k]}$. Assume that $\overline{\mathcal{V}} \subset \mathcal{V}$ is a set of bad vertices. Set also $\mathcal{V}^* := \mathcal{V} \setminus \overline{\mathcal{V}}$. Let $\overline{L}$ be the free $\mathbb{Z}$–submodule of L spanned by the (base elements of) $\overline{\mathcal{V}}$, and let $\overline{L}_{\geq 0}$ be its first quadrant. Next we introduce a special weight function for the points of $\overline{L}_{\geq 0}$ (which, in general, is not a Riemann–Roch type formula, it is not even quadratic).

For any $\mathbf{i} = (i_{v_1}, \dots) \in \overline{L}_{\geq 0}$ we define the element $x(\mathbf{i}) \in L$ by the following universal property:

(i) the coefficient of E_v in $x(\mathbf{i})$ is i_v for any $v \in \overline{\mathcal{V}}$,
(ii) $(x(\mathbf{i}) + s_{[k]}, E_v) \leq$ for every $v \in \mathcal{V}^*$,
(iii) $x(\mathbf{i})$ is minimal with the properties (i)-(ii).

The definition is motivated by the theory of generalized computation sequences used in singularity theory Laufer [1972, 1977] and Némethi [1999a,b].

Set $w_{k_r}(\mathbf{i}) := \chi_{k_r}(x(\mathbf{i}))$, the weight function of $\overline{L}_{\geq 0}$. Using this weight function, one defines the weight of any cube of $\overline{L}_{\geq 0}$ as the maximum of the weights of the vertices of the cube, and one also repeats the definition of S_n and of the lattice cohomology similarly as above.

Theorem 5.8.2. Reduction Theorem. *László and Némethi [2015]* $\mathbb{H}^*(L_{\geq 0}, \chi_{k_r}) = \mathbb{H}^*(\overline{L}_{\geq 0}, w_{k_r})$.

In particular, if Γ is AR (and $\mathbb{H}^p = 0$ for any $p > 0$ by Theorem 5.4.1) $\mathbb{H}^0(\Gamma, [k])$ for any $spin^c$–structure $[k]$ is determined by the sequence of integers $\{\tau(i)\}_{i \in \mathbb{Z}_{\geq 0}}$, $\tau(i) := \chi_{k_r}(x(i))$, where i is the coordinate of the bad vertex v. This was intensively used in concrete computations. The star–shaped graphs are AR, their τ–functions are described in terms of Seifert invariants in Némethi [2005].

If $\mathcal{K} \subset S^3$ is an algebraic knot (that is, one of its representative can be cut out by an isolated irreducible plane curve singularity germ $f_{\mathcal{K}}$), then the surgery 3–manifold $S^3_{-p/q}(\mathcal{K})$ for $p/q \in \mathbb{Q}_{>0}$ is a plumbed 3–manifold associated with a negative definite AR graph. For their τ–function see Némethi [2007]. This can be determined in terms of the semigroup of $f_{\mathcal{K}}$.

More generally, if $\{\mathcal{K}_i\}_{i=1}^N$ are algebraic knots then the graph of $S^3_{-p/q}(\#_i \mathcal{K})$ and the reduced weights together with the lattice cohomology in terms of the semigroups are determined in Némethi and Román [2012].

Theorem 3.2.1 has its 'reduction' as well. Indeed, one defines the reduction $Z^{\overline{\mathcal{V}}}$ of the series $Z(\mathbf{t})$ to the variables $\{t_v\}_{v \in \overline{\mathcal{V}}}$ (similarly as in Sections 2.5 and 3.3, by taking $t_v = 1$ for any $v \notin \overline{\mathcal{V}}$), and László and Némethi [2015] proves for it the analogue of Theorem 3.2.1. In particular, the Seiberg–Witten invariants can also be recovered as the periodic constants of the reduces series.

5.9 Ehrhart theory. In László and Némethi [2014] the Seiberg–Witten invariant (of a negative definite plumbed 3–manifold) is identified with the third coefficient of certain equivariant Ehrhart polynomial.

5.10 $Z(\mathbf{t})$ and $\mathfrak{eu}$ in terms of weighted cubes. Above in the definition of the lattice cohomology we used for each $spin^c$–structure a different weighted lattice. This can be unified in a common weighted lattice Némethi [2011] and Ozsváth, Stipsicz, and Szabó [2014]. Here we follow Némethi [2011]. The set of p–cubes consists of pairs (k, I), where $k \in \mathrm{Char}$ and $I \subset \mathcal{V}$, $|I| = p$. This pair can be identified with a cube in $L \otimes \mathbb{R}$ with vertices $\{k + 2E_{I'}\}_{I' \subset I}$. One defines the weight function induced by the intersection form $w : \mathrm{Char} \to \mathbb{Q}$ by $w(k) := -(k^2 + |\mathcal{V}|)/8$, which extends to a weight–function of the q–cubes via $w((k, I)) = \max_{I' \subset I}\{ w(k + 2E_{I'})\}$. (The correspondence between the two languages, for any fixed $[k]$, is realized by $k = K + 2s_{[k]} + 2l \leftrightarrow l$ and a universal constant shift in the weights.) The point is that in this language of weighted cubes the topological series $Z(\mathbf{t})$ can also be expressed Némethi [ibid.] as:

$$Z(\mathbf{t}) = \sum_{k \in \mathrm{Char}} \sum_{I \subset \mathcal{V}} (-1)^{|I|+1}\, w((k, I)) \cdot \mathbf{t}^{\frac{1}{2}(k-K)}. \tag{5.10.1}$$

Note that $\mathrm{Char} = K + 2L' \subset L'$, hence $(k - K)/2$ runs over L' when k runs over Char.

The normalized Seiberg–Witten invariant, or the normalized Euler characteristic of the lattice cohomology can also be determined directly by a weighted cube counting (this is really the analogue how Euler defined the classical Euler characteristic via alternating sum of p–cells/cubes). Since we have infinitely many cubes in $L \otimes \mathbb{R}$, we need to consider a 'truncation'.

Let us fix the class $[k]$. For any two $k_1, k_2 \in [k] \subset \mathrm{Char}$ with $k_1 \le k_2$ let $R(k_1, k_2)$ denote the rectangle $\{x \in L \otimes \mathbb{R} \ : \ k_1 \le x \le k_2\}$. Then we can consider the set of cubes (k, I) with $k \in [k]$ and situated with all vertices in $R(k_1, k_2)$ and the corresponding lattice cohomology $\mathbb{H}^*(R(k_1, k_2), w)$.

Theorem 5.10.2. *For 'good' choices of k_1, k_2 (with $k_1 \ll 0$ and $k_2 \gg 0$), and with the abridgment $R = R(k_1, k_2)$, one has the isomorphism of $\mathbb{Z}[U]$–modules $\mathbb{H}^*(\Gamma, k_r) = \mathbb{H}^*(R, w)$. Furthermore,*

$$\begin{aligned} \mathfrak{eu}(\mathbb{H}^*(\Gamma, k_r)) &= -\min(w|R) + \sum_p (-1)^p \,\mathrm{rank}_{\mathbb{Z}}\ \mathbb{H}^p_{red}(R, w) \\ &= \sum_{(k,I) \subset R} (-1)^{|I|+1} w((k, I)). \end{aligned}$$

Note that $Z(\mathbf{t})$ determines the intersection form $(\ ,\ )$ of the lattice L, hence the lattice cohomology as well (note also the direct 'periodic constant formula' for the Seiberg–Witten invariant).

Problem 5.10.3. Find a universal construction, which assigns to any series $S(\mathbf{t})$ a graded $\mathbb{Z}[U]$–module, such that it applied to $Z(\mathbf{t})$ we recover $\mathbb{H}^*$.

5.11 The fundamental group of M. In fact, the same question can be asked for the fundamental group of M. Note that $\pi_1(M)$ determines M in a unique way (except for lens spaces, a case which can be disregarded, since they have trivial $\mathbb{H}^*_{red}$).

Problem 5.11.1. Find a universal construction, which assigns to any group a graded $\mathbb{Z}[U]$–module, such that it applied to $\pi_1(M)$ we recover $\mathbb{H}^*$.

5.12 Another surgery formula for $\mathbb{H}^*$. Recall the surgery formulae 3.3.1 valid for the Seiberg–Witten invariant involving the periodic constant as a 'correction term'.

Problem 5.12.1. Find its analogue at the level of $\mathbb{H}^*$. (This is related with Problem 5.10.3 too).

5.13 Path lattice cohomology. Consider the situation of Section 4.2 with fixed Γ, and $k = K$ and weight function $\chi = \chi_K$. Furthermore, consider a sequence $\gamma := \{x_i\}_{i=0}^t$ so that $x_0 = 0$, $x_{i+1} = x_i + E_{v(i)}$ for certain $v(i) \in \mathcal{V}$ for each $0 \leq i < t$, and $x(t) \in -K + \mathcal{S}'$. Then for the union of 0–cubes marked by the points $\{x_i\}_i$ and segments (1–cubes) of type $[x_i, x_{i+1}]$ we can repeat the definition of the lattice cohomology (associated with χ_k), and we get a graded $\mathbb{Z}[U]$–module $\mathbb{H}^*(\gamma, K)$. In fact, only $\mathbb{H}^0$ is nonzero, and $\mathbb{H}^0(\gamma, K) = \mathcal{T}^+_{2\min_i\{\chi(x_i)\}} \oplus \mathbb{H}^0_{red}(\gamma, K)$. This is called the path–cohomology associated with the 'path' γ and χ. Similarly as in Section 5.6, we consider its normalized Euler characteristic

$$eu(\mathbb{H}^0(\gamma, K)) := -\min_i\{\chi(x_i)\} + \mathrm{rank}_{\mathbb{Z}}\, \mathbb{H}^0_{red}(\gamma, K).$$

One shows that

$$eu(\mathbb{H}^0(\gamma, K)) = \sum_{i=0}^{t-1} \max\{\chi(x_i), \chi(x_{i+1})\} - \chi(x_{i+1}) = \sum_{i=0}^{t-1} \max\{(x_i, E_{v(i)}) - 1, 0\}.$$

It is convenient to introdce the following notation as well (cf. Section 5.6):

$$eu(\mathbb{H}^0(\Gamma, K)) := -\min\chi + \mathrm{rank}\, \mathbb{H}^0_{red}(\Gamma, K).$$

It turns out (cf. Némethi and Sigurdsson [2016]) that

$$\min_\gamma\, eu(\mathbb{H}^0(\gamma, K)) \leq eu(\mathbb{H}^0(\Gamma, K)).$$

5.14 Graded roots. For each Γ and $k \in \mathrm{Char}$, the author in Némethi [2005] constructed a *graded root*, from which one recovers by a natural procedure $\mathbb{H}^0(\Gamma, k)$. This is tree whose vertices are $\mathbb{Z}$–graded, the weight n–vertices correspond to the connected components of S_n, and the edges of the tree to the possible inclusions of the components of S_n to the components of S_{n+1} (this at the $\mathbb{H}^0(\Gamma, k)$ level is codified in the U–action). For AR graphs it contains all the lattice cohomology (hence Heegaard–Floer) information, but it codifies completely the inclusions of the S_n–components (a fact, which from the U–action cannot be recovered).

There is a parallel characterization of rational and elliptic graphs in terms of graded roots too.

5.15 Classification of singularities. The ‘Artin-Laufer program’ starts the classification of singularities with the class of rational and elliptic singularities — it identifies topologically the families, and then (under certain analytic conditions in the elliptic case, e.g. Gorenstein property) determines certain analytic invariants (multiplicity, Hilbert function) by uniform formulae for each family.

In order to continue this program, we first have to identify subfamilies for which one can show that they share ‘common properties’. We propose to identify these subfamilies by graded roots, or by the (slightly weaker) $\mathbb{Z}[U]$–module $\mathbb{H}^0(\Gamma, K)$ (or, we can even consider the whole $\mathbb{H}^*(\Gamma, K)$) Némethi [2007].

Though we know all the possible topological types (namely, the singularity resolution graphs are exactly the connected negative definite graphs), it is not clear at all what graded tree might appear as a graded root associated with a resolution graph.

Problem 5.15.1. Classify all the possible graded roots associated with all (Γ, K) of normal surface singularities. Is there any hidden structure property carried by the $\mathbb{Z}[U]$–modules $\mathbb{H}^*(\Gamma, k)$ associated with topological type of singularities? It is possible to describe all the possible modules produced in this way?

6 Analytic – topological connections. The Seiberg–Witten Invariant Conjecture

In this section we treat a set of possible properties connecting the analytic invariants with the topological ones, namely, the equivariant geometric genera with the Seiberg–Witten invariants of the link. When they are valid they provide a topological description of the equivariant geometric genera. The identities are generalizations of the statement of the Casson Invariant Conjecture of Neumann and Wahl to the case of normal surface singularities with rational homology sphere links.

Conjecture 6.0.1. Seiberg–Witten Invariant Conjecture/Coincidence.

We say that (X,o) satisfy the equivariant SWIC if for any $h \in H$ the following identity holds

$$h^1(\widetilde{X}, \mathcal{O}(-r_h)) = -\mathfrak{sw}_{-h*\sigma_{can}}(M) - \frac{(K+2r_h)^2+|\mathcal{V}|}{8}. \tag{6.0.2}$$

Its validity automatically extends to arbitrary natural line bundles as follows:

$$h^1(\widetilde{X}, \mathcal{O}(l')) = -\mathfrak{sw}_{[l']*\sigma_{can}}(M) - \frac{(K-2l')^2+|\mathcal{V}|}{8}. \tag{6.0.3}$$

We say that (X,o) satisfies the SWIC if the above identity holds for $h=0$, that is, if

$$p_g(X,o) = -\mathfrak{sw}_{\sigma_{can}}(M) - \frac{K^2+|\mathcal{V}|}{8}. \tag{6.0.4}$$

The identity SWIC was formulated as a conjecture in Némethi and Nicolaescu [2002], while the equivariant case in Némethi [2007]: the expectation was that it holds for any $\mathbb{Q}$–Gorenstein singularity. Now we know that this is not true for this large class of singularities (see Luengo-Velasco, Melle-Hernández, and Némethi [2005]), although it is valid for large number of smaller families of singularities. But even in the case of those families when it fails, it still indicates interesting 'virtual' properties. The limits of the validity of the properties are not clarified at this moment. Having in mind the existence of cases when the identity does not hold, one might say that it is not totally justified the name SWI 'Conjecture', although this was its name in the literature in the last ten years. Hence, the reader might read the abbreviation SWIC as SWI 'Coincidence' too.

Example 6.0.5. CIC of Neumann and Wahl [1990] Assume that (X,o) is Gorenstein and it admits a smoothing with smooth nearby Milnor fiber F. Then the signature of F satisfies $\sigma(F)+8p_g+K^2+|\mathcal{V}|=0$, hence the SWIC for $h=0$ reads as $\sigma(F)/8 = \mathfrak{sw}_{\sigma_{can}}(M)$. (In this case, usually, $\sigma(F)/8$ is not an integer.) Additionally, if (X,o) is a complete intersection with integral homology sphere link, then $\mathfrak{sw}_{\sigma_{can}}(M)$ equals the Casson invariant $\lambda(M)$ of M, hence the above identity reads as $\sigma(F)/8=\lambda(M)$. This is the Casson Invariant Conjecture of Neumann and Wahl, predicted for any complete intersection with integral homology sphere link Neumann and Wahl [ibid.].

The CIC was solved for weighted homogeneous singularities and hypersurface suspension singularities ($(X,o)=\{f(x,y)+z^N=0\}$) in Neumann and Wahl [ibid.], for splice singularities (which includes the weighted homogeneous case as well) in Némethi and Okuma [2009]. The general case is still open.

6.1 Regarding the validity of SWIC.

We have the following statement:

Theorem 6.1.1. *The equivariant SWIC was verified in the following cases: rational singularities Némethi and Nicolaescu [2002], weighted homogeneous singularities Némethi and Nicolaescu [2002, 2004], splice quotient singularities Némethi and Okuma [2008].*

Additionally, the SWIC (for $h = 0$) was verified for suspensions $\{f(x, y) + z^N = 0\}$ with f irreducible Némethi and Nicolaescu [2005], for hypersurface Newton non–degenerate singularities Sigurdsson [2016] and superisolated singularities with one cusp Fernández de Bobadilla, Luengo-Velasco, Melle-Hernández, and Némethi [2006, 2007] and Borodzik and Livingston [2014]. Since the identity of the SWIC is stable with respect to equisingular deformations, the SWIC remains valid for such deformations of any of the above cases.

Recall that (X, o) is a hypersurface superisolated singularity if

$$(X, o) = \{f(x_1, x_2, x_3) = 0\}$$

where f is a hypersurface singularity with isolated singularity and the homogeneous terms $f_d + f_{d+1} + \cdots$ of f satisfy the following properties: $C := \{f_d = 0\}$ is reduced and it defines in $\mathbb{CP}^2$ an irreducible rational cuspidal curve C; furthermore, the intersection $\{f_{d+1} = 0\} \cap \mathrm{Sing}\{f_d = 0\}$ in $\mathbb{CP}^2$ is empty. The restriction regarding f_d implies that the link of (X, o) is a rational homology sphere. One shows that the minimal good graph of (X, o) has ν bad vertices, where ν is the number of cusps of C.

In all cases $p_g = d(d-1)(d-2)/6$, hence it depends only on d, however the Seiberg–Witten invariant (and the plumbing graph too) depend essentially on the type of cusps of C (see Section 6.4 below).

For superisolated singularities in certain cases when C is not unicuspidal the SWIC ($h = 0$ case) is not true Luengo-Velasco, Melle-Hernández, and Némethi [2005].

6.2 The path lattice cohomology bound of p_g. The failure of the SWIC in the case of superisolated singularities motivated a parallel deeper study of these germs. Surprisingly, in this case a rather natural universal bound of p_g will become equality.

Consider the set of paths $\mathfrak{P}$, $\gamma := \{x_i\}_{i=0}^t$ so that $x_0 = 0$, $x_{i+1} = x_i + E_{v(i)}$ for certain $v(i) \in \mathcal{V}$ for each $0 \leq i < t$, and $x(t) \in -K + \mathcal{S}'$. (If Γ is numerically Gorenstein we can even take $x(t) = -K$.) Then, considering the cohomology exact sequences $0 \to \mathcal{O}_{E_{v(i)}}(-x_i) \to \mathcal{O}_{x_{i+1}} \to \mathcal{O}_{x_i} \to 0$, we obtain that $p_g \leq \mathrm{eu}(\mathbb{H}^0(\gamma, K))$. Therefore,

$$p_g \leq \min_{\gamma \in \mathfrak{P}} \mathrm{eu}(\mathbb{H}^0(\gamma, K)). \tag{6.2.1}$$

This inequality is in the spirit of computational sequences initiated by Laufer and used intensively in Laufer [1972, 1977], Némethi [1999a], and László and Némethi [2015]. From this point of view, this relation, in some sense, is even more natural then that required by the SWIC. Note that in Equation (6.2.1) equality holds if and only if all the cohomology exact sequence (along a path which minimalizes the right hand side) split. The point is that this happens for certain key families.

Theorem 6.2.2 (Némethi and Sigurdsson [2016]). *The identity*

$$p_g = \min_{\gamma \in \mathfrak{P}} \mathfrak{eu}(\mathbb{H}^0(\gamma), K)$$

holds in the following cases:

(a) if $p_g = \mathfrak{eu}(\mathbb{H}^0(\Gamma, K))$ *(this happens e.g. if* $\mathbb{H}^q(\Gamma, K) = 0$ *for* $q \geq 1$ *and* (X, o) *satisfies the SWIC). In particular,* $p_g = \min_{\gamma \in \mathfrak{P}} \mathfrak{eu}(\mathbb{H}^0(\gamma), K)$ *holds for all weighted homogeneous and minimally elliptic singularities.*

(b) rational and Gorenstein elliptic singularities;

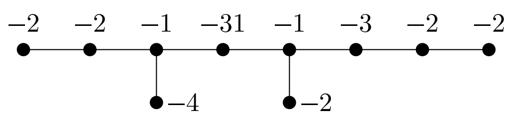
(c) for superisolated singularities (with arbitrary number of cusps);

(d) for hypersurface singularities with non-degenerate Newton principal part.

Again, since the identity is stable with respect to equisingular deformations, it remains valid for such deformations of any of the above cases.

The next example shows that in the case of superisolated singularities the non-vanishing of $\mathbb{H}^q(M)$ $(1 \leq q < \nu)$ obstructs the validity of the SWIC.

Example 6.2.3. Némethi and Sigurdsson [ibid.] Assume that (X, o) is a superisolated singularity with C of degree $d = 5$ and two cusps, both with one Puiseux pair: $(3, 4)$ and $(2, 7)$ respectively. The graph Γ is



We set $k = K$. One shows that $\min \chi = -5$, $\mathrm{rank}_{\mathbb{Z}}(\mathbb{H}^0_{red}) = 5$, $\mathrm{rank}_{\mathbb{Z}}(\mathbb{H}^1) = 2$. Hence $\mathfrak{eu}(\mathbb{H}^*) = 8$. Since for the superisolated germ with $d = 5$ one has $p_g = 10$, using Section 5.6 we get that the SWIC is not valid. On the other hand, $\min_\gamma \mathfrak{eu}(\mathbb{H}^0(\gamma, K)) = 10$ as well, hence $p_g = \min_\gamma \mathfrak{eu}(\mathbb{H}^0(\gamma, K))$, as it is predicted by the above theorem.

If we take any other analytic structure supported by Γ, by Equation (6.2.1) $p_g \leq 10$ still holds.

This graph (that is, topological type) supports another natural analytic structure as well, namely a splice quotient analytic type (compare with Section 7.2): it is the $\mathbb{Z}_5$–factor of the complete intersection $\{z_1^3 + z_2^4 + z_3^5 z_4 = z_3^7 + z_4^2 + z_1^4 z_2 = 0\} \subset (\mathbb{C}^4, 0)$ by the diagonal action $(\alpha^2, \alpha^4, \alpha, \alpha)$ $(\alpha^5 = 1)$. By Theorem 6.1.1 it satisfies the SWIC, hence $p_g = 8$.

In particular, in their choices of the topological characterization of their geometric genus, some analytic structures prefer $eu(\mathbb{H}^*(\Gamma, K))$, some of them the extremal

$$\min_{\gamma} eu(\mathbb{H}^0(\gamma, K))$$

(and there might exists even other choices, as parts/versions of the lattice cohomology). From this point of view the abridgement SWIC might also mean 'SWI Choice'.

6.3 What is the optimal topological lower/upper bounds of p_g? We are guided by the following key question: If one fixes a topological type (say, a minimal good resolution graph) and varies the possible analytic structures supported on this fixed topological type, then what are the possible values of the geometric genus p_g? A more concrete version is formulated as follows:

Problem 6.3.1. Associate combinatorially a concrete integer $\mathrm{MAX}(\Gamma)$ to any resolution graph Γ, such that for any analytic type supported by Γ one has $p_g \leq \mathrm{MAX}(\Gamma)$, and furthermore, for certain analytic structure one has equality.

Obviously, one can ask for the symmetric $\mathrm{MIN}(\Gamma)$ as well. But for the optimal lower bound we know the answer. A possible candidate is the 'artihmetical genus' $p_a(X, o) = 1 - \min \chi$ Wagreich [1970]. Indeed, for any analytic structure, whenever $p_g > 0$, one also has $1 - \min\chi \leq p_g$ Wagreich [ibid., p. 425].

The inequality looks not very sharp, however we have the following statement:

Theorem 6.3.2. *For any non–rational topological type $p_g = 1 - \min\chi$ for the generic analytic structure. (In particular, for non–rational graphs $\mathrm{MIN}(\Gamma)$ is exactly $1 - \min\chi$.)*

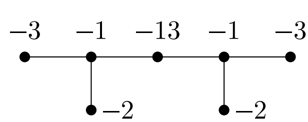
This was proved for elliptic singularities in Laufer [1977], the general case in Némethi and Nagy [n.d.].

A possible upper bound for p_g, hence a candidate for $\mathrm{MAX}(\Gamma)$, is $\min_{\gamma\in\mathfrak{P}} eu(\mathbb{H}^0(\gamma))$, cf. Equation (6.2.1).

However, for the next graph, $p_g \leq \min_{\gamma\in\mathfrak{P}} eu(\mathbb{H}^0(\gamma))$ is not sharp. Indeed

$$\min_{\gamma\in\mathfrak{P}} eu(\mathbb{H}^0(\gamma)) = 4$$

while for *any* analytic type $p_g \leq 3$ Némethi and Okuma [2017]. (For any Gorenstein structure one has $p_g = 3$, nevertheless, $p_g = 3$ can be realized by non–Gorenstein structure as well.)



6.4 Superisolated singularities revisited. Assume that the superisolated germ f is associated with the projective rational cuspidal curve C, as in Section 6.1. Here two main parts of the algebraic/analytic geometry meat: the classification of projective plane curves with the theory of local singularities. One of the aims of the classification of projective plane curves is to list the set of local topological types of plane curve singularities which can be realized as singularities of a degree d projective cuspidal curve. The strategy is to impose restrictions, obstructions for such realizations from the theory of singularities applied for the corresponding superisolated singularities. E.g., if $\{\mathcal{K}_i \subset S^3\}_{i=1}^{\nu}$ are the local algebraic knots of the plane curve singularities $\{(C,p_i)\}_{i=1}^{\nu}$ of C, and $\deg(C)=d$, then the link of the superisolated germ (X,o) is $M = S^2_{-d}(\#_i\mathcal{K}_i)$. Furthermore, if μ_i is the Milnor number of $(\mathcal{K}_i \subset S^3)$, then by the genus formula and the rationality of C implies $\sum_i \mu_i = (d-1)(d-2)$, hence d is uniquely determined by the local knot–types. The question whether we can impose any other restriction on M from the existence of C.

In the next presentation we follow Bodnár and Némethi [2016] and Fernández de Bobadilla, Luengo-Velasco, Melle-Hernández, and Némethi [2006, 2007]. Let us introduce the following notations. For each (C,p_i) (or $\mathcal{K}_i \subset S^3$) let Δ_i denote the Alexander polynomial (normalized as $\Delta_i(1)=1$) and $\Gamma_i \subset \mathbb{Z}_{\geq 0}$ the semigroup of (C,p_i). Recall that $\Delta_i(t) = (1-t)\cdot\sum_{k\in\Gamma_i} t^k$ Guseĭn-Zade, Delgado, and Kampil'o [1999], and $\#\{\mathbb{Z}_{\geq 0}\setminus\Gamma_i\} = \delta_i = \mu_i/2$ (the so-called delta–invariant, or genus of (C,p_i)), and write $\delta := \sum_i \delta_i = (d-1)(d-2)/2$.

Furthermore, consider the product of Alexander polynomials:

$$\Delta(t) := \Delta_1(t)\Delta_2(t)\cdots\Delta_\nu(t).$$

There is a unique polynomial Q for which $\Delta(t) = 1+\delta(t-1)+(t-1)^2Q(t)$. Write $Q(t) = \sum_{j=0}^{2\delta-2} q_j t^j$. For $\nu=1$ one shows

$$Q(t) = \sum_{s\notin\Gamma_1}(1+t+\cdots+t^{s-1}), \text{ hence } q_j = \#\{s\notin\Gamma_1 : s>j\} \quad (\text{if } \nu=1). \tag{6.4.1}$$

Next, set the rational function

$$R(t) := \frac{1}{d}\sum_{\xi^d=1}\frac{\Delta(\xi t^{1/d})}{(1-\xi t^{1/d})^2} - \frac{1-t^d}{(1-t)^3}. \tag{6.4.2}$$

In Fernández de Bobadilla, Luengo-Velasco, Melle-Hernández, and Némethi [2006, (2.4)] is proved that $R(t)$ is a symmetric polynomial ($R(t) = t^{d-3}R(1/t)$), and

$$R(t) = \sum_{j=0}^{d-3}\Big(q_{(d-3-j)d} - \frac{(j+1)(j+2)}{2}\Big)t^{d-3-j}. \tag{6.4.3}$$

In fact, if we reduce the variables of $Z(\mathbf{t})$ and $\mathcal{P}(\mathbf{t})$ to the variable t corresponding to $\mathfrak{U} := \{C\} \subset \mathcal{V}$ then $R(t) = Z^{\mathfrak{U}}_{h=0}(t) - \mathcal{P}^{\mathfrak{U}}_{h=0}(t)$, hence, by surgery formulae 2.6.1 and 3.3.1 we get that

$$(6.4.4) \qquad R(1) = -\mathfrak{sw}_{\sigma_{can}}(M) - (K^2 + |\mathcal{V}|)/8 - p_g(X,o).$$

Hence the vanishing of $R(1)$ is equivalent with the SWIC for (X,o). Moreover, in Fernández de Bobadilla, Luengo-Velasco, Melle-Hernández, and Némethi [ibid.] is also proved that for $\nu = 1$ all the coefficients of $R(t)$ are nonnegative, hence $(R(1) \leq 0) \Leftrightarrow (R(1) = 0) \Leftrightarrow$ all coefficients of $R(t)$ are zero. On the other hand, by Theorem 6.1.1, for $\nu = 1$ one has $R(1) = 0$ indeed. This implies the vanishing of all coefficients of $R(t)$; hence Equations (6.4.1) and (6.4.3) combined provide the strong *semigroup distribution property* of Γ_1, which must be satisfied by any collection of local knot types and degree d whenever the data is realized by a curve C.

In Fernández de Bobadilla, Luengo-Velasco, Melle-Hernández, and Némethi [ibid.] we conjectured that for any rational cuspidal plane curve $C \subset \mathbb{C}P^2$ of degree d with arbitrary number of cusps all the coefficients of $R(t)$ are non–positive. In Bodnár and Némethi [2016] is proved that this is indeed true for any $\nu \leq 2$, but it fails, in general, for $\nu \geq 3$. The corrected conjecture, as formulated in Bodnár and Némethi [ibid.] is the following:

Conjecture 6.4.5. Bodnár and Némethi [ibid.] For any superisolated germ

$$R(1) \leq 0, \quad \text{that is,} \quad p_g \geq -\mathfrak{sw}_{\sigma_{can}}(M) - (K^2 + |\mathcal{V}|)/8.$$

In Bodnár and Némethi [ibid.] this conjecture is proved for $\nu \leq 2$ and verified for all 'known' rational cuspidal curves with $\nu \geq 3$. The next reformulation of this conjecture in terms of the lattice cohomology emphasize once again the differences and resemblances between $\mathfrak{eu}(\mathbb{H}^*(\Gamma, K))$ and $\mathfrak{eu}(\mathbb{H}^0(\Gamma, K))$.

Conjecture 6.4.6. *Bodnár and Némethi [ibid.] For the link $M = S^3_{-d}(\#_i \mathcal{K}_i)$ of a superisolated surface singularity corresponding to a rational cuspidal projective plane curve of degree d we have:*

$$\mathfrak{eu}(\mathbb{H}^*(M,K)) \leq \mathfrak{eu}(\mathbb{H}^0(M,K)).$$

In fact, $\mathfrak{eu}(\mathbb{H}^0(M,K)) = d(d-1)(d-2)/6$ by Borodzik and Livingston [2014] and Bodnár and Némethi [2016]. From this reformulation it is clear its validity for $\nu \leq 2$, however a conceptual argument for its validity for $\nu \geq 3$ is still missing.

6.5 Coverings. Let (X,o) be a normal surface singularity with rational homology sphere link, and let (Y,o) be its universal abelian covering, cf. Section 2.1. Then, from Equation (2.1.2) automatically one has the analytic identity

$$p_g(Y,o) = \sum_{h\in H} h^1(\widetilde{X}, \mathcal{O}_{\widetilde{X}}(-r_h)), \tag{6.5.1}$$

that is, the sum of the equivariant geometric genera of (X,o) equals the geometric genus of (Y,o).

Let M and N be the links of (X,o) and (Y,o) respectively. Then N is the regular universal abelian covering of M (associated with the representation $\pi_1(M)\to H$). Let $\Gamma(M)$ and $\Gamma(N)$ be the corresponding negative definite resolution/plumbing graphs as well. In the next discussion we also assume that N is a rational homology sphere too.

Having in mind the equivariant SWIC identities for (X,o) and the SWIC for (Y,o), following Bodnár and Némethi [2017] we say that M and its universal abelian covering N satisfy the '*covering additivity property*' (CAP) if

$$\mathfrak{sw}_{\sigma_{can}}(N) + \frac{K(\Gamma(N))^2 + |\mathcal{V}(\Gamma(N))|}{8} =$$
$$= \sum_{h\in H} \mathfrak{sw}_{-h*\sigma_{can}}(M) + \frac{(K(\Gamma(M)) + 2r_h)^2 + |\mathcal{V}(\Gamma(N))|}{8}.$$

Clearly, if (X,o) satisfies the equivariant SWIC and (Y,o) the SWIC, then (CAP) for M and its universal abelian covering N holds. However, since (CAP) is a totally topological identity, we might ask its validity for any singularity link M (whose universal abelian covering N is a rational homology sphere), independently of the existence of any nice analytic structure, or independently of singularity theory.

The point is that the property (CAP) in general is not true Bodnár and Némethi [ibid.]. But, what is really surprising is that (CAP) is true for the surgery 3–manifolds $M = S^3_{-p/q}(\#_i \mathcal{K}_i)$, even though (some of) these 3–manifolds appear as the links of superisolated singularities, and the superisolated singularities are the basic couterexamples for SWIC.

Theorem 6.5.2. *Bodnár and Némethi [ibid.] Let $M = S^3_{-p/q}(K)$ be a manifold obtained by a negative rational Dehn surgery of S^3 along a connected sum of algebraic knots $\mathcal{K} = \mathcal{K}_1\#\dots\#\mathcal{K}_\nu$ ($p,q>0$, $\gcd(p,q)=1$). Assume that N, the universal abelian covering of M, is a rational homology sphere. Then (CAP) holds.*

Both statements Equation (6.5.1) and Theorem 6.5.2 remain valid for any abelian covering.

Problem 6.5.3. Characterize those 3–manifolds M (or, singularity links) for which (CAP) holds.

6.6 Newton non–degenerate germs revisited. Note that for a hypersurface Newton non–degenerate isolated singularity (X, o), with rational homology sphere link, both identities are true: $p_g = \mathfrak{eu}(\mathbb{H}^*(\Gamma, K))$ by Theorem 6.1.1, and

$$p_g = \min_{\gamma \in \mathfrak{P}} \mathfrak{eu}(\mathbb{H}^0(\gamma, K))$$

by Theorem 6.2.2. In particular, the link of such a singularity satisfy a very strong topological restiction: $\mathfrak{eu}(\mathbb{H}^*(\Gamma, K)) = \min_{\gamma \in \mathfrak{P}} \mathfrak{eu}(\mathbb{H}^0(\gamma, K))$. Though we know that such 3–manifolds are rather special (for the algorithm which determines the plumbing graph from the Newton diagram see Oka [1987, 1997], or Braun and Némethi [2010]), still, this topological identity is a mystery for us. (Maybe it is worth to mention that in this case, p_g also equals the lattice points of $(\mathbb{Z}_{>0})^3$ below the Newton diagram.)

Finally, we end this section with the following

Problem 6.6.1. Find the Heegaard–Floer theoretical interpretation of

$$\min_{\gamma \in \mathfrak{P}} \mathfrak{eu}(\mathbb{H}^0(\gamma, K))$$

7 Analytic – topological connections. $\mathcal{P}(\mathbf{t})$ versus $Z(\mathbf{t})$

7.1 The $\mathcal{P}(\mathbf{t}) = Z(\mathbf{t})$ identity. Recall that by Section 2.4 the periodic constant of $\mathcal{P}_h$ is the equivarint geometric genus $h^1(\widetilde{X}, \mathcal{O}(-r_h))$, while by Theorem 3.2.1 the periodic constant of Z_h is the opposite of the r_h–normalized Seiberg–Witten invariant. Hence the equivariant SWIC says that the periodic constants of $\mathcal{P}_h$ and Z_h are equal. Hence, it is natural to ask for the validity of an even stronger identity, namely for $\mathcal{P}(\mathbf{t}) = Z(\mathbf{t})$. If this identity holds, then it provides a topological characterisation of the multivariable Hilbert function of the divisorial filtration. For some simple singularities, e.g. for cyclic quotient singularities one can compute directly both sides, and their equality follows visibly.

Theorem 7.1.1. *The equality $\mathcal{P}(\mathbf{t}) = Z(\mathbf{t})$ is true in the following cases: (a) rational singularities Campillo, Delgado, and Gusein-Zade [2004] and Guseĭn-Zade, Delgado, and Kampil'o [2008] (for a different proof see Némethi [2008b]); (b) minimally elliptic singularities Némethi [ibid.]; (c) splice quotient singularities Némethi [2012] (this includes e.g. the weighted homogeneous case as well).*

Even the 'reduced identity' $\mathcal{P}^{\mathcal{U}} = Z^{\mathcal{U}}$, for some subset $\mathcal{U} \subset \mathcal{V}$, can be interesting, even if $\mathcal{U}$ contains only one element v. In the splice quotient case $\mathcal{P}^{\{v\}} = Z^{\{v\}}$ for certain nodes was proved and used in Okuma [2008]. In the weighted homogeneous case, when the minimal good graph is star–shaped and v is the central vertex, then $\mathcal{P}^{\{v\}}_{h=0}$ coincides also with the Poincaré series of the $\mathbb{Z}_{\geq 0}$–graded algebra of the singular point (grading induced by the $\mathbb{C}^*$–action). For this series Pinkham provided a topological expression in terms of the Seifert invariants of the Seifert 3–manifold link Pinkham [1977], which coincides with $Z^{\{v\}}_{h=0}$ (see e.g. Némethi and Nicolaescu [2004]). In the case of suspension singularities the difference $\mathcal{P}^{\{C\}}_{h=0} - Z^{\{C\}}_{h=0}$ is captured by the polynomial R, see Section 6.4.

We note also that the identity $P_h = Z_h$ is much stronger than SWIC for h: one can construct examples when $P_h \neq Z_h$ but the SWIC holds. On the other hand, in any situation when the SWIC fails, the identity $\mathcal{P} = Z$ also fails (e.g. for certain superisolated singularities, see also Némethi [2008b]).

7.2 A closer look at splice quotient singularities. Splice quotient singularities were introduced by Neumann and Wahl [2005b,a]. From any fixed graph Γ (whose plumbed manifold is a rational homology sphere and which has some additional special arithmetical properties, see below) one constructs a family of singularities with common equisingularity type, such that any member admits a distinguished resolution, whose dual graph is exactly Γ. The construction suggests that the analytic properties of the singularities constructed in this way are strongly linked with a fixed resolution and with its graph Γ. (Hence, the expectation is that certain analytic invariants might be computable from Γ.)

In present, there are three different approaches how one can introduce and study splice quotient singularities; each of them is based on a different geometric property. They are: (a) the 'original' construction of Neumann–Wahl, (b) the 'modified' version by Okuma [2008], and (c) considering singularities satisfying the 'end–curve condition'. It turns out that all these approaches are equivalent.

In the first two cases we start with a topological type (that is, with Γ), which satisfies certain restrictions, and we endow it with an analytic structure, the 'splice quotient' analytic type. In the third case we start with an analytic structure, which satisfies a certain analytic property.

The construction of Neumann and Wahl [2005a] imposes two combinatorial restriction on Γ, the *semigroup and congruence conditions*. The congruence condition is empty if $\det(,) = 1$. Using the first condition one writes the equations of a complete intersection (Y, o). The equations depend only on the splice diagram associated with the graph, in particular they are called 'splice diagram equations'. Then one defines an action of H on this complete intersection, free off o, (here the congruence condition is needed), and sets $(X, o) = (Y, o)/G$. It turns out that (X, o) has a resolution with dual graph Γ and (Y, o) is the universal abelian covering of (X, o).

Okuma replaces the semigroup and congruence conditions by the *monomial conditions*, otherwise the construction and the output is the same. Singularities constructed in this way (either Neumann–Wahl version or Okuma version) are called splice quotient singularities.

The third approach defines a family of singularities with a special analytic property, with the *end curve condition*. This requires the existence of a resolution which has the following property: For each exceptional irreducible component E_v, which corresponds to an end–vertex of Γ, there exists an analytic function whose reduced strict transform is irreducible, it intersects the exceptional curve only along E_v, and this intersection is transversal. These are called ‘end curve functions’.

Theorem 7.2.1. *(I) (Topological part) Fix a graph* Γ. *The following facts are equivalent:*

(a) There exists a splice quotient singularity with resolution graph Γ*,*

(b) Γ *satisfies the semigroup and congruence conditions;*

(c) Γ *satisfies the monomial condition.*

(II) (Analytic part) Fix a normal surface singularity (X, o). *The following facts are equivalent:*

(a) (X, o) *is splice quotient (in the sense of Neumann–Wahl or Okuma);*

(b) (X, o) *satisfies the end curve condition;*

(c) $\mathcal{P}(\mathbf{t}) = Z(\mathbf{t})$.

(I) follows from Neumann and Wahl [ibid.], the equivalence $(I)(b) \Leftrightarrow (c)$ was proved in Neumann and Wahl [ibid., §13]). Regarding (II), the fact that splice quotient singularities satisfy the ‘end curve condition’ follows basically from the construction of the singularities: some powers of the coordinate functions of (Y, o) are ‘end curve function’. The converse is the subject of the ‘End Curve Theorem’ Neumann and Wahl [2010] and Okuma [2010]. Part $(b) \Rightarrow (c)$ was proved in Némethi [2012], part $(c) \Rightarrow (b)$ follows from definitions.

Example 7.2.2. The end curve condition is satisfied in the following cases: (a) rational singularities, where π is an arbitrary resolution; (b) minimally elliptic singularities, and π is a minimal resolution; (c) weighted homogeneous singularities, where π is the minimal good resolution.

As we already said, we do not know the ‘limits’ of the SWIC, but the stronger version $\mathcal{P}(\mathbf{t}) = Z(\mathbf{t})$ occurs exactly when (X, o) is splice quotient (provided that M is a rational homology sphere). In this case the analytic $\mathcal{P}$ make the topological choice Z.

Problem 7.2.3. Find other topological candidates for $\mathcal{P}(\mathbf{t})$ — or, transformed into a question: what other topological choices might have $\mathcal{P}(\mathbf{t})$ when we vary the analytic structure of (X, o) ?

What is the ‘multivariable series lift’ of $\min_{\gamma \in \mathfrak{P}} \mathfrak{eu}(\mathbb{H}^0(\gamma, K))$? What is $\mathcal{P}_{h=0}(\mathbf{t})$ for hypersurface superisolated or Newton non–degenerate germs ? (In the last cases it can even happen that $\mathcal{P}_{h=0}(\mathbf{t})$ is not constant along the equisingular strata, then find $\mathcal{P}$ for some ‘normal form’.) Describe/characterize the universal abelian covers of hypersurface superisolated or Newton non–degenerate singularities.

7.3 The analytic semigroup of (X,o). In Section 2.1 we introduced the 'topological semigroup' $\mathcal{S}'$ determined from the lattice L'. Its analytic counterpart is defined as

$$\mathcal{S}'_{an,h} = \{r_h + \mathrm{div}_E(s), \text{ where } s \in H^0(\mathcal{O}_{\widetilde{X}}(-r_h))\}$$

for any $h \in H$, and $\mathcal{S}'_{an} = \cup_h \mathcal{S}'_{an,h}$. One shows that $\mathcal{S}'_{an} \subset \mathcal{S}'$, and $\mathcal{P}(\mathbf{t})$ is supported in $\mathcal{S}'_{an}$, while $Z(\mathbf{t})$ is supported in $\mathcal{S}'$. In both cases, usually the supports are much smaller then the corresponding semigroups. However, both semigroups conceptually guide important geometric properties of the analytic/topological type of (X,o). For rational or minimally elliptic singularities $\mathcal{S}'_{an} = \mathcal{S}'$, however the computation of $\mathcal{S}'_{an}$ usually is extremely hard.

Problem 7.3.1. Find efficient methods for the computation of $\mathcal{S}'_{an}$. Describe $\mathcal{S}'_{an}$ for certain key families of singularities.

7.4 Linear subspace arrangements, as 'lifts' of the series $\mathcal{P}(\mathbf{t})$ and $Z(\mathbf{t})$. Fix (X,o), a resolution π and the filtration $\{\mathcal{F}(l')\}_{l'\in L'}$ as in Section 2.2. For any $l' \in L'$, the linear space

$$(\mathcal{F}(l')/\mathcal{F}(l'+E))_{\theta([l'])} = H^0(\mathcal{O}_{\widetilde{X}}(-l'))/H^0(\mathcal{O}_{\widetilde{X}}(-l'-E))$$

naturally embeds into

$$T(l') := H^0(\mathcal{O}_E(-l')).$$

Let its image be denoted by $A(l')$. Furthermore, for every $v \in \mathcal{V}$, consider the linear subspace $T_v(l')$ of $T(l')$ given by

$$T_v(l') := H^0(\mathcal{O}_{E-E_v}(-l'-E_v)) = \ker\big(\, H^0(\mathcal{O}_E(-l')) \to H^0(\mathcal{O}_{E_v}(-l'))\,\big) \subset T(l').$$

Then the image $A_v(l')$ of $H^0(\mathcal{O}_{\widetilde{X}}(-l'-E_v)/H^0(\mathcal{O}_{\widetilde{X}}(-l'-E))$ in $T(l')$ satisfies $A_v(l') = A(l') \cap T_v(l')$.

The point is that one can show that the vector space $T(l')$ and the linear subspace arrangement $\{T_v(l')\}_v$ in $T(l')$ depends only on the resolution graph.

Definition 7.4.1. The (finite dimensional) arrangement of linear subspaces $\mathcal{A}_{\mathrm{top}}(l') = \{T_v(l')\}_v$ in $T(l')$ is called the 'topological arrangement' at $l' \in L'$. The arrangement of linear subspaces $\mathcal{A}_{\mathrm{an}}(l') = \{A_v(l') = T_v(l') \cap A(l')\}_v$ in $A(l')$ is called the 'analytic arrangement' at $l' \in L'$. The corresponding projectivized arrangement complements will be denoted by $\mathbb{P}(T(l') \setminus \cup_v T_v(l'))$ and $\mathbb{P}(A(l') \setminus \cup_v A_v(l'))$ respectively.

If $l' \notin \mathcal{S}'$ then there exists v such that $(E_v, l') > 0$, that is $h^0(\mathcal{O}_{E_v}(-l')) = 0$, proving that $T_v(l') = T(l')$. Hence $A_v(l') = A(l')$ too. In particular, both arrangement complements are empty. In fact, if $l' \notin \mathcal{S}'_{an}$, then by similar argument, the analytic arrangement complement is empty too.

The connection with the series $\mathcal{P}(\mathbf{t})$ and $\mathbb{Z}(\mathbf{t})$ is given by the following topological Euler characteristic formulae associated with all the linear subspace arrangements for all l'.

Theorem 7.4.2.

$$\mathcal{P}(\mathbf{t}) = \sum_{l' \in \mathcal{S}'_{an}} \chi_{\text{top}}(\,\mathbb{P}(A(l') \setminus \cup_v A_v(l'))\,) \cdot \mathbf{t}^{l'};$$

$$Z(\mathbf{t}) = \sum_{l' \in \mathcal{S}'} \chi_{\text{top}}(\,\mathbb{P}(T(l') \setminus \cup_v T_v(l'))\,) \cdot \mathbf{t}^{l'}.$$

For proof see Némethi [n.d.(a)], the analytic case (in the language of $\{\mathcal{F}(l')\}_{l'}$) already appeared in Campillo, Delgado, and Gusein-Zade [2004] and Guseĭn-Zade, Delgado, and Kampil'o [2008].

The corresponding dimensions of the linear subspaces in $\mathcal{Q}_{\text{an}}(l')$ are as follows. For any $l' \in L'$ and $I \subset \mathcal{V}$ one has $\dim A(l') = \mathfrak{h}(l' + E) - \mathfrak{h}(l')$, $\dim \cap_{v \in I} A_v(l') = \mathfrak{h}(l' + E) - \mathfrak{h}(l' + E_I)$. Thus, the analytic arrangement is rather sensitive to the modification of the analytic structure, and in general, does not coincide with the topological arrangement. The corresponding dimensions in the topological case are computed in Némethi [n.d.(a)], they are slightly more technical. Examples show that $\chi_{\text{top}}(\mathbb{P}(A(l') \setminus \cup_v A_v(l'))) = \chi_{\text{top}}(\mathbb{P}(T(l') \setminus \cup_v T_v(l')))$ can happen even if $A(l') \neq T(l')$.

7.4.3. Note that the analytic subspace arrangement $\mathcal{Q}_{an}(l')$, naturally determined by the divisorial filtration, exists even without its embedding into $T(l')$. On the other hand, if one wishes to find its topological analogue, its embedding into the topological $T(l')$ is the most natural possibility, that is, the choice of $\mathcal{Q}_{top}(l')$ is the most natural universal object, which might include all the possible analytic arrangements indexed by different analytic structures.

In this way, $(A(l'), \{A_v(l')\}_v) \subset (T(l'), \{T_v(l')\}_v)$ looks a perfect pairing. This immediately induces (by taking the Euler characteristic of the corresponding spaces) the two series $Z(\mathbf{t})$ and $\mathcal{P}(\mathbf{t})$. Though these two series looked artificially paired at the beginning, now, after Theorem 7.4.2, this fact is totally motivated and justified. Furthermore, taking the periodic constants of the series Z and P, we get that the pairing predicted by the SWIC is indeed very natural and totally justified.

In particular, these steps provide a totally conceptual explanation for the appearance of the Seiberg–Witten invariant in the theory of complex surface singularities.

7.4.4. Extensions. The above picture allows to extend the series $\mathcal{P}(\mathbf{t})$ and $Z(\mathbf{t})$ to capture some additional information from the corresponding Hodge or Grothendieck ring structures as well. In the analytic case the extension of $\mathcal{P}(\mathbf{t})$ to the series $\sum_{l' \in \mathcal{S}'} [\mathbb{P}(A(l') \setminus$

$\cup_v A_v(l'))] \cdot \mathbf{t}^{l'}$ with coefficients in the Grothendieck ring was already considered e.g. in Campillo, Delgado, and Gusein-Zade [2007]. Once we have the topological arrangement in hand, we can also introduce the following series with coefficients in the Grothendieck ring

$$Z(\mathbb{L},\mathbf{t}) = \sum_{l'\in\mathcal{S}'} [\mathbb{P}(T(l') \setminus \cup_v T_v(l'))] \cdot \mathbf{t}^{l'}. \tag{7.4.5}$$

It is remarkable that this series has a closed expression in terms of lattice too. Indeed, if $\mathcal{E}$ denotes the set of edges of Γ, then (see Némethi [n.d.(a)])

$$Z(\mathbb{L},\mathbf{t}) = \frac{\prod_{(u,v)\in\mathcal{E}} (1-\mathbf{t}^{E_u^*} - \mathbf{t}^{E_v^*} + \mathbb{L}\mathbf{t}^{E_u^*+E_v^*})}{\prod_{v\in\mathcal{V}} (1-\mathbf{t}^{E_v^*})(1-\mathbb{L}\mathbf{t}^{E_v^*})}. \tag{7.4.6}$$

References

Michael Artin (1962). "Some numerical criteria for contractability of curves on algebraic surfaces". *Amer. J. Math.* 84, pp. 485–496. MR: 0146182 (cit. on p. 777).

– (1966). "On isolated rational singularities of surfaces". *Amer. J. Math.* 88, pp. 129–136. MR: 0199191 (cit. on p. 777).

József Bodnár and András Némethi (2016). "Lattice cohomology and rational cuspidal curves". *Math. Res. Lett.* 23.2, pp. 339–375. MR: 3512889 (cit. on pp. 788, 789).

– (2017). "Seiberg-Witten invariant of the universal abelian cover of $S^3_{-p/q}(K)$". In: *Singularities and computer algebra*. Springer, Cham, pp. 173–197. MR: 3675727 (cit. on p. 790).

Maciej Borodzik and Charles Livingston (2014). "Heegaard Floer homology and rational cuspidal curves". *Forum Math. Sigma* 2, e28, 23. MR: 3347955 (cit. on pp. 785, 789).

Steven Boyer, Cameron McA. Gordon, and Liam Watson (2013). "On L-spaces and left-orderable fundamental groups". *Math. Ann.* 356.4, pp. 1213–1245. MR: 3072799 (cit. on p. 779).

Gábor Braun and András Némethi (2010). "Surgery formula for Seiberg-Witten invariants of negative definite plumbed 3-manifolds". *J. Reine Angew. Math.* 638, pp. 189–208. MR: 2595340 (cit. on pp. 770, 773, 774, 791).

A. Campillo, F. Delgado, and S. M. Gusein-Zade (2004). "Poincaré series of a rational surface singularity". *Invent. Math.* 155.1, pp. 41–53. MR: 2025300 (cit. on pp. 770, 791, 795).

– (2007). "Multi-index filtrations and generalized Poincaré series". *Monatsh. Math.* 150.3, pp. 193–209. MR: 2308548 (cit. on p. 796).

Steven Dale Cutkosky, Jürgen Herzog, and Ana Reguera (2004). "Poincaré series of resolutions of surface singularities". *Trans. Amer. Math. Soc.* 356.5, pp. 1833–1874. MR: 2031043 (cit. on p. 770).

J. Fernández de Bobadilla, I. Luengo-Velasco, A. Melle-Hernández, and A. Némethi (2006). "On rational cuspidal projective plane curves". *Proc. London Math. Soc. (3)* 92.1, pp. 99–138. MR: 2192386 (cit. on pp. 785, 788, 789).

– (2007). "On rational cuspidal plane curves, open surfaces and local singularities". In: *Singularity theory*. World Sci. Publ., Hackensack, NJ, pp. 411–442. MR: 2342921 (cit. on pp. 785, 788).

Javier Fernández de Bobadilla, Ignacio Luengo-Velasco, Alejandro Melle Hernández, and András Némethi (2007). "Classification of rational unicuspidal projective curves whose singularities have one Puiseux pair". In: *Real and complex singularities*. Trends Math. Birkhäuser, Basel, pp. 31–45. MR: 2280130.

Joshua Evan Greene (2013). "A surgery triangle for lattice cohomology". *Algebr. Geom. Topol.* 13.1, pp. 441–451. MR: 3031647 (cit. on pp. 774, 776).

S. M. Guseĭn-Zade, F. Delgado, and A. Kampil'o (1999). "On the monodromy of a plane curve singularity and the Poincaré series of its ring of functions". *Funktsional. Anal. i Prilozhen.* 33.1, pp. 66–68. MR: 1711890 (cit. on p. 788).

– (2008). "Universal abelian covers of rational surface singularities, and multi-index filtrations". *Funktsional. Anal. i Prilozhen.* 42.2, pp. 3–10, 95. MR: 2438013 (cit. on pp. 770, 791, 795).

Jonathan Hanselman, Jacob Rasmussen, Sarah Dean Rasmussen, and Liam Watson (n.d.). "Taut foliations on graph manifolds". arXiv: 1508.05911 (cit. on p. 779).

Tamás László, János Nagy, and András Némethi (n.d.). "Surgery formulae for the Seiberg-Witten invariant of plumbed 3-manifolds". arXiv: 1702.06692v1 (cit. on p. 774).

Tamás László and András Némethi (2014). "Ehrhart theory of polytopes and Seiberg-Witten invariants of plumbed 3-manifolds". *Geom. Topol.* 18.2, pp. 717–778. MR: 3180484 (cit. on pp. 770, 771, 780).

– (2015). "Reduction theorem for lattice cohomology". *Int. Math. Res. Not. IMRN* 11, pp. 2938–2985. MR: 3373041 (cit. on pp. 779, 780, 785).

Henry B. Laufer (1971). *Normal two-dimensional singularities*. Annals of Mathematics Studies, No. 71. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, pp. xi+161. MR: 0320365.

– (1972). "On rational singularities". *Amer. J. Math.* 94, pp. 597–608. MR: 0330500 (cit. on pp. 777, 780, 785).

– (1977). "On minimally elliptic singularities". *Amer. J. Math.* 99.6, pp. 1257–1295. MR: 0568898 (cit. on pp. 775, 777, 780, 785, 787).

Christine Lescop (1996). *Global surgery formula for the Casson-Walker invariant*. Vol. 140. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. ii+150. MR: 1372947 (cit. on p. 773).

Yuhan Lim (2000). "Seiberg-Witten invariants for 3-manifolds in the case $b_1 = 0$ or 1". *Pacific J. Math.* 195.1, pp. 179–204. MR: 1781619 (cit. on p. 773).

I. Luengo-Velasco, A. Melle-Hernández, and A. Némethi (2005). "Links and analytic invariants of superisolated singularities". *J. Algebraic Geom.* 14.3, pp. 543–565. MR: 2129010 (cit. on pp. 784, 785).

Ignacio Luengo-Velasco (1987). "The μ-constant stratum is not smooth". *Invent. Math.* 90.1, pp. 139–152. MR: 906582.

András Némethi (n.d.[a]). Book in preparation (cit. on pp. 795, 796).

– (n.d.[b]). "Line bundles associated with normal surface singularities". arXiv: math.AG/0310084 (cit. on p. 769).

– (n.d.[c]). "On the Heegaard Floer homology of $S^3_{-p/q}(K)$". arXiv: math.GT/0410570.

– (1999a). ""Weakly" elliptic Gorenstein singularities of surfaces". *Invent. Math.* 137.1, pp. 145–167. MR: 1703331 (cit. on pp. 775, 780, 785).

– (1999b). "Five lectures on normal surface singularities". In: *Low dimensional topology (Eger, 1996/Budapest, 1998)*. Vol. 8. Bolyai Soc. Math. Stud. With the assistance of Ágnes Szilárd and Sándor Kovács. János Bolyai Math. Soc., Budapest, pp. 269–351. MR: 1747271 (cit. on pp. 775, 777, 780).

– (2005). "On the Ozsváth-Szabó invariant of negative definite plumbed 3-manifolds". *Geom. Topol.* 9, pp. 991–1042. MR: 2140997 (cit. on pp. 776–780, 782).

– (2007). "Graded roots and singularities". In: *Singularities in geometry and topology*. World Sci. Publ., Hackensack, NJ, pp. 394–463. MR: 2311495 (cit. on pp. 780, 783, 784).

– (2008a). "Lattice cohomology of normal surface singularities". *Publ. Res. Inst. Math. Sci.* 44.2, pp. 507–543. MR: 2426357 (cit. on pp. 775, 777–779).

– (2008b). "Poincaré series associated with surface singularities". In: *Singularities I*. Vol. 474. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 271–297. MR: 2454352 (cit. on pp. 769, 791, 792).

– (2011). "The Seiberg-Witten invariants of negative definite plumbed 3-manifolds". *J. Eur. Math. Soc. (JEMS)* 13.4, pp. 959–974. MR: 2800481 (cit. on pp. 773, 779, 781).

András Némethi (2011). "Two exact sequences for lattice cohomology". In: *Proceedings of the conference organized to honor H. Moscovici's 65th birthday, Contemporary Math*. Vol. 546, pp. 249–269 (cit. on pp. 774, 776, 778).

András Némethi (2012). "The cohomology of line bundles of splice-quotient singularities". *Adv. Math.* 229.4, pp. 2503–2524. MR: 2880230 (cit. on pp. 791, 793).

– (2017). "Links of rational singularities, L-spaces and LO fundamental groups". *Invent. Math.* 210.1, pp. 69–83. MR: 3698339 (cit. on p. 779).

András Némethi and János Nagy (n.d.). Manuscript in preparation (cit. on p. 787).

András Némethi and Liviu I. Nicolaescu (2002). "Seiberg-Witten invariants and surface singularities". *Geom. Topol.* 6, pp. 269–328. MR: 1914570 (cit. on pp. 773, 784, 785).

– (2004). "Seiberg-Witten invariants and surface singularities. II. Singularities with good $\mathbb{C}^*$-action". *J. London Math. Soc. (2)* 69.3, pp. 593–607. MR: 2050035 (cit. on pp. 785, 792).

– (2005). "Seiberg-Witten invariants and surface singularities: splicings and cyclic covers". *Selecta Math. (N.S.)* 11.3-4, pp. 399–451. MR: 2215260 (cit. on p. 785).

András Némethi and Tomohiro Okuma (2008). "The Seiberg-Witten invariant conjecture for splice-quotients". *J. Lond. Math. Soc. (2)* 78.1, pp. 143–154. MR: 2427056 (cit. on p. 785).

– (2009). "On the Casson invariant conjecture of Neumann-Wahl". *J. Algebraic Geom.* 18.1, pp. 135–149. MR: 2448281 (cit. on pp. 770, 784).

András Némethi and Tomohiro Okuma (2017). "Analytic singularities supported by a specific integral homology sphere link, Methods and Applications of Analysis". In: Special volume dedicated to Henry Laufer's 70th birthday on February 15, 2017 (Conference at Sanya, China) (cit. on p. 787).

András Némethi and Fernando Román (2012). "The lattice cohomology of $S^3_{-d}(K)$". In: *Zeta functions in algebra and geometry*. Vol. 566. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 261–292. MR: 2858927 (cit. on p. 780).

András Némethi and Baldur Sigurdsson (2016). "The geometric genus of hypersurface singularities". *J. Eur. Math. Soc. (JEMS)* 18.4, pp. 825–851. MR: 3474458 (cit. on pp. 782, 786).

Walter D. Neumann and Jonathan Wahl (1990). "Casson invariant of links of singularities". *Comment. Math. Helv.* 65.1, pp. 58–78. MR: 1036128 (cit. on p. 784).

– (2002). "Universal abelian covers of surface singularities". In: *Trends in singularities*. Trends Math. Birkhäuser, Basel, pp. 181–190. MR: 1900786.

– (2003). "Universal abelian covers of quotient-cusps". *Math. Ann.* 326.1, pp. 75–93. MR: 1981612.

– (2005a). "Complete intersection singularities of splice type as universal abelian covers". *Geom. Topol.* 9, pp. 699–755. MR: 2140991 (cit. on pp. 792, 793).

– (2005b). "Complex surface singularities with integral homology sphere links". *Geom. Topol.* 9, pp. 757–811. MR: 2140992 (cit. on p. 792).

– (2010). "The end curve theorem for normal complex surface singularities". *J. Eur. Math. Soc. (JEMS)* 12.2, pp. 471–503. MR: 2608949 (cit. on p. 793).

Liviu I. Nicolaescu (2004). "Seiberg-Witten invariants of rational homology 3-spheres". *Commun. Contemp. Math.* 6.6, pp. 833–866. MR: 2111431 (cit. on p. 773).

Mutsuo Oka (1987). "On the resolution of the three-dimensional Brieskorn singularities". In: *Complex analytic singularities*. Vol. 8. Adv. Stud. Pure Math. North-Holland, Amsterdam, pp. 437–460. MR: 894304 (cit. on p. 791).

– (1997). *Non-degenerate complete intersection singularity*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, pp. viii+309. MR: 1483897 (cit. on p. 791).

Tomohiro Okuma (2004). "Universal abelian covers of rational surface singularities". *J. London Math. Soc. (2)* 70.2, pp. 307–324. MR: 2078895.

– (2006). "Universal abelian covers of certain surface singularities". *Math. Ann.* 334.4, pp. 753–773. MR: 2209255.

– (2008). "The geometric genus of splice-quotient singularities". *Trans. Amer. Math. Soc.* 360.12, pp. 6643–6659. MR: 2434304 (cit. on pp. 769, 770, 772, 791, 792).

– (2010). "Another proof of the end curve theorem for normal surface singularities". *J. Math. Soc. Japan* 62.1, pp. 1–11. MR: 2648226 (cit. on p. 793).

Peter S Ozsváth and Zoltán Szabó (n.d.). "Knot Floer homology and rational surgeries". arXiv: math.GT/0504404.

Peter Ozsváth, András I. Stipsicz, and Zoltán Szabó (2014). "A spectral sequence on lattice homology". *Quantum Topol.* 5.4, pp. 487–521. MR: 3317341 (cit. on pp. 778, 781).

Peter Ozsváth and Zoltán Szabó (2003a). "Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary". *Adv. Math.* 173.2, pp. 179–261. MR: 1957829.

– (2003b). "On the Floer homology of plumbed three-manifolds". *Geom. Topol.* 7, pp. 185–224. MR: 1988284 (cit. on pp. 775, 776).

– (2004a). "Holomorphic disks and topological invariants for closed three-manifolds". *Ann. of Math. (2)* 159.3, pp. 1027–1158. MR: 2113019 (cit. on p. 774).

– (2004b). "Holomorphic triangle invariants and the topology of symplectic four-manifolds". *Duke Math. J.* 121.1, pp. 1–34. MR: 2031164.

H. Pinkham (1977). "Normal surface singularities with C^* action". *Math. Ann.* 227.2, pp. 183–193. MR: 0432636 (cit. on p. 792).

Baldur Sigurdsson (2016). "The geometric genus and Seiberg-Witten invariant of Newton nondegenerate surface singularities". PhD Thesis, CEU Budapest, Hungary. PhD thesis. arXiv: 1601.02572 (cit. on p. 785).

András Szenes and Michèle Vergne (2003). "Residue formulae for vector partitions and Euler-MacLaurin sums". *Adv. in Appl. Math.* 30.1-2. Formal power series and algebraic combinatorics (Scottsdale, AZ, 2001), pp. 295–342. MR: 1979797 (cit. on p. 771).

Vladimir Turaev (Jan. 2001). "Surgery formula for torsions and Seiberg–Witten invariants of 3-manifolds". arXiv: math/0101108 (cit. on p. 774).

Philip Wagreich (1970). "Elliptic singularities of surfaces". *Amer. J. Math.* 92, pp. 419–454. MR: 0291170 (cit. on pp. 777, 787).

Received 2017-12-05.

András Némethi
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Reáltanoda utca 13-15
H-1053, Budapest
Hungary

and

ELTE - University of Budapest
Dept. of Geometry
Budapest
Hungary

and

BCAM - Basque Center for Applied Math., Mazarredo
14 E48009 Bilbao
Basque Country
Spain
nemethi.andras@renyi.mta.hu

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (799–824)

$\mathscr{D}$-MODULES IN BIRATIONAL GEOMETRY

MIHNEA POPA

Abstract

It is well known that numerical quantities arising from the theory of $\mathscr{D}$-modules are related to invariants of singularities in birational geometry. This paper surveys a deeper relationship between the two areas, where the numerical connections are enhanced to sheaf theoretic constructions facilitated by the theory of mixed Hodge modules. The emphasis is placed on the recent theory of Hodge ideals.

1 Introduction

Ad hoc arguments based on differentiating rational functions or sections of line bundles abound in complex and birational geometry. To pick just a couple of examples, topics where such arguments have made a deep impact are the study of adjoint linear series on smooth projective varieties, see for instance Demailly's work on effective very ampleness Demailly [1993] and its more algebraic incarnation in Ein, Lazarsfeld, and Nakamaye [1996], and the study of hyperbolicity, see for instance Siu's survey Siu [2004] and the references therein.

A systematic approach, as well as an enlargement of the class of objects to which differentiation techniques apply, is provided by the theory of $\mathscr{D}$-modules, which has however only recently begun to have a stronger impact in birational geometry. The new developments are mainly due to a better understanding of Morihiko Saito's theory of mixed Hodge modules Saito [1988], Saito [1990], and hence to deeper connections with Hodge theory and coherent sheaf theory. Placing problems in this context automatically brings in important tools such as vanishing theorems, perverse sheaves, or the V-filtration, in a unified way.

Connections between invariants arising from log resolutions of singularities and invariants arising from the theory of $\mathscr{D}$-modules go back a while however. A well-known such

The author was partially supported by the NSF grant DMS-1700819.
MSC2010: primary 14F10; secondary 14J17, 32S25, 14F17, 14F18, 14C30.

instance is the fact that the log canonical threshold of a function f on (say) $\mathbb{C}^n$ coincides with the negative of the largest root of the Bernstein-Sato polynomial $b_f(s)$; see e.g. Yano [1983], Kollár [1997]. Numerical data on log resolutions plays a role towards the study of other roots of the Bernstein-Sato polynomial Kashiwara [1976/77], Lichtin [1989], though our understanding of these is far less thorough. Going one step further, a direct relationship between the multiplier ideals of a hypersurface in a smooth variety X and the V-filtration it induces on $\mathscr{O}_X$ was established in Budur and Saito [2005].

After reviewing some of this material, in this paper I focus on one direction of further development, worked out jointly with Mustaţă Mustaţă and Popa [2016a], Mustaţă and Popa [2016b], Mustaţă and Popa [2018a], Mustaţă and Popa [2018b] as well as by Saito in Saito [2016], namely the theory of what we call *Hodge ideals*. This is a way of thinking about the Hodge filtration (in the sense of mixed Hodge modules) on the sheaf of functions with arbitrary poles along a hypersurface, or twists thereof, and is closely related to both the singularities of the hypersurface and the Hodge theory of its complement. There are two key approaches that have proved useful towards understanding Hodge ideals:

1. A birational study in terms of log resolutions, modeled on the algebraic theory of multiplier ideals, which Hodge ideals generalize, Mustaţă and Popa [2016a], Mustaţă and Popa [2018a].

2. A comparison with the (microlocal) V-filtration, using its interaction with the Hodge filtration in the case of mixed Hodge modules, Saito [2016], Mustaţă and Popa [2018b].

Hodge ideals are indexed by the non-negative integers; at the 0-th step, they essentially coincide with multiplier ideals. Beyond the material presented in this paper, by analogy it will be interesting to develop a theory of Hodge ideals associated to ideal sheaves (perhaps leading to asymptotic constructions as well), to attempt an alternative analytic approach, and to establish connections with constructions in positive characteristic generalizing test ideals.

There are other ways in which filtered $\mathscr{D}$-modules underlying Hodge modules have been used in recent years in complex and birational geometry, for instance in the study of generic vanishing theorems, holomorphic forms, topological invariants, families of varieties and hyperbolicity; see e.g. Dimca, Maisonobe, and Saito [2011], Schnell [2012], Popa and Schnell [2013], Wang [2016], Popa and Schnell [2014], Popa and Schnell [2017], Pareschi, Popa, and Schnell [2017], Wei [2017]. The bulk of the recent survey Popa [2016b] treats part of this body of work, so I have decided not to discuss it here again. In any event, the reader is advised to use Popa [ibid.] as a companion to this article, as introductory material on $\mathscr{D}$-modules and Hodge modules together with a guide to technical literature can be found there (especially in Ch. B, C). Much of that will not be repeated here, for space reasons.

Acknowledgments. Most of the material in this paper describes joint work with Mircea Mustaţă, and many ideas are due to him. Special thanks are due to Christian Schnell, who helped my understanding of Hodge modules through numerous discussions and collaborations. I am also indebted to Morihiko Saito, whose work and ideas bear a deep influence on the topics discussed here. Finally, I thank Yajnaseni Dutta, Mircea Mustaţă, Lei Wu and Mingyi Zhang for comments on a first draft.

2 V-filtration, Bernstein-Sato polynomial, and birational invariants

One of the main tools in the theory of mixed Hodge modules is the V-filtration along a hypersurface, and its interaction with the Hodge filtration. Important references regarding the V-filtration include Kashiwara [1983], Malgrange [1983], Sabbah [1987], Saito [1988], Saito [1994].

First, let's recall the graph construction. Let D be an effective divisor on X, given (locally, in coordinates $x_1, \ldots, x_n$) by $f = 0$ with $f \in \mathscr{O}_X$, and whose support is Z. Consider the embedding of X given by the graph of f, namely:

$$i_f = (\mathrm{id}, f)\colon X \hookrightarrow X \times \mathbb{C} = Y, \quad x \to \big(x, f(x)\big).$$

On $\mathbb{C}$ we consider the coordinate t, and a vector field ∂_t such that $[\partial_t, t] = 1$.

Let $(\mathscr{M}, F)$ be a filtered left $\mathscr{D}_X$-module. We denote

$$(\mathscr{M}_f, F) := i_{f_+}(\mathscr{M}, F) = (\mathscr{M}, F) \otimes_{\mathbb{C}} (\mathbb{C}[\partial_t], F),$$

where the last equality (which means the filtration is the convolution filtration) is the definition of push-forward for filtered $\mathscr{D}$-modules via a closed embedding. More precisely, we have

- $\mathscr{M}_f = \mathscr{M} \otimes_{\mathbb{C}} \mathbb{C}[\partial_t]$, with action of $\mathscr{D}_Y = \mathscr{D}_X[t, \partial_t]$ given by: $\mathscr{O}_X$ acts by functions on $\mathscr{M}$, and
$$\partial_{x_i} \cdot (g \otimes \partial_t^i) = \partial_{x_i} g \otimes t^i - (\partial_{x_i} f) g \otimes \partial_t^{i+1},$$
$$t \cdot (g \otimes \partial_t^i) = fg \otimes \partial_t^i - i g \otimes \partial_t^{i-1}, \quad \text{and} \quad \partial_t \cdot (g \otimes \partial_t^i) = g \otimes \partial_t^{i+1}.$$

- $F_p \mathscr{M}_f = \bigoplus_{i=0}^p F_{p-i} \mathscr{M} \otimes \partial_t^i$ for all $p \in \mathbb{Z}$.

One of the key technical tools in the study of $\mathscr{D}$-modules is the V-filtration. The $\mathbb{Z}$-indexed version always exists and is unique when $\mathscr{M}_f$ is a regular holonomic $\mathscr{D}_Y$-module, due to work of Kashiwara [1983] and Malgrange [1983]. Assuming in addition that the local monodromy along f is quasi-unipotent, a condition of Hodge-theoretic origin satisfied by all the objects appearing here, one can also consider the following $\mathbb{Q}$-indexed version; cf. Saito [1988, p. 3.1.1].

Definition 2.1 (Rational V-filtration). A *rational V-filtration* of $\mathscr{M}_f$ is a decreasing filtration $V^\alpha \mathscr{M}_f$ with $\alpha \in \mathbb{Q}$ satisfying the following properties:

- The filtration is exhaustive, i.e. $\bigcup_\alpha V^\alpha \mathscr{M}_f = \mathscr{M}_f$, and each $V^\alpha \mathscr{M}_f$ is a coherent $\mathscr{O}_Y[\partial_{x_i}, \partial_t t]$-submodule of $\mathscr{M}_f$.
- $t \cdot V^\alpha \mathscr{M}_f \subseteq V^{\alpha+1}\mathscr{M}_f$ and $\partial_t \cdot V^\alpha \mathscr{M}_f \subseteq V^{\alpha-1}\mathscr{M}_f$ for all $\alpha \in \mathbb{Q}$.
- $t \cdot V^\alpha \mathscr{M}_f = V^{\alpha+1}\mathscr{M}_f$ for $\alpha > 0$.
- The action of $\partial_t t - \alpha$ on $\mathrm{gr}_V^\alpha \mathscr{M}_f$ is nilpotent for each α. (One defines $\mathrm{gr}_\alpha^V \mathscr{M}_f$ as $V^\alpha \mathscr{M}_f / V^{>\alpha}\mathscr{M}_f$, where $V^{>\alpha}\mathscr{M}_f = \cup_{\beta>\alpha} V^\beta \mathscr{M}_f$.)

We will consider other $\mathscr{D}$-modules later on, but for the moment let's focus on the case $\mathscr{M} = \mathscr{O}_X$, with the trivial filtration $F_k\mathscr{O}_X = \mathscr{O}_X$ for $k \geq 0$, and $F_k\mathscr{O}_X = 0$ for $k < 0$. It is standard to denote $\mathscr{B}_f := (\mathscr{O}_X)_f$. Via the natural inclusion of $\mathscr{O}_X$ in $\mathscr{B}_f$, for $\alpha \in \mathbb{Q}$ one defines

$$V^\alpha \mathscr{O}_X := V^\alpha \mathscr{B}_f \cap \mathscr{O}_X,$$

a decreasing sequence of coherent ideal sheaves on X. A first instance of the connections we focus on here is the following result of Budur-Saito:

Theorem 2.2 (Budur and Saito [2005, Theorem 0.1]). *If D is an effective divisor on X, then for every $\alpha \in \mathbb{Q}$ one has*

$$V^\alpha \mathscr{O}_X = \mathscr{J}\big((\alpha - \epsilon)D\big),$$

the multiplier ideal of the $\mathbb{Q}$-divisor $(\alpha-\epsilon)D$, where $0 < \epsilon \ll 1$ is a rational number.

Multiplier ideal sheaves are ubiquitous objects in birational geometry, encoding local numerical invariants of singularities, and satisfying Kodaira-type vanishing theorems in the global setting; see Lazarsfeld [2004, Ch. 9]. If $f : Y \to X$ is a log resolution of the pair (X, D), and $c \in \mathbb{Q}$, then by definition the multiplier ideal of cD is

$$\mathscr{J}(cD) = f_*\mathscr{O}_Y\big(K_{Y/X} - [cf^*D]\big).$$

Let me take the opportunity to also introduce the following notation, to be used repeatedly. Denote $Z = D_{\mathrm{red}}$, and define integers a_i, b_i and c_i by the expressions

$$f^*Z = \tilde{Z} + a_1F_1 + \cdots + a_mF_m$$

and

$$K_{Y/X} = b_1 F_1 + \cdots + b_m F_m + c_{m+1} F_{m+1} + \cdots + c_n F_n,$$

where F_j are the components of the exceptional locus and $a_i \neq 0$. We denote

$$\gamma = \min_{1 \leq i \leq m} \frac{b_i + 1}{a_i}. \tag{2.3}$$

Recall that the *Bernstein-Sato polynomial* of f is the unique monic polynomial $b_f(s)$ of minimal degree, in the variable s, such that there exists $P \in \mathscr{D}_X[s]$ satisfying the formal identity

$$b_f(s) f^s = P f^{s+1}.$$

See for instance Kashiwara [1976/77], Sabbah [1987], Saito [1994], while a nice survey can be found in Granger [2010]. It can be shown that $b_f(s)$ is independent of the choice of f such that $D = \mathrm{div}(f)$ locally, and so one also has a function $b_D(s)$ which is globally well defined; however, to keep a unique simple notation, in the statements below all information about the pair (X, D) related to $b_f(s)$ should be understood locally in this sense.

The roots of the Bernstein-Sato polynomial are interesting invariants of the singularities of f, and a number of important facts regarding them have been established in the literature. Here are some of the most significant; a posteriori, many of these facts also follow from Theorem 2.2 and the connection between the Bernstein-Sato polynomial and the V-filtration.

1. The roots of $b_f(s)$ are negative rational numbers; see Kashiwara [1976/77].
2. More precisely, in the notation above, all the roots of $b_f(s)$ are of the form $-\frac{b_i + \ell}{a_i}$ for some $i \geq 0$ and $\ell \geq 1$; see Lichtin [1989, Theorem 5].
3. The negative α_f of the largest root of $b_f(s)$ is the log canonical threshold of (X, D); Kollár [1997, Theorem 10.6], see also Yano [1983], Lichtin [1989].
4. Moreover, all jumping numbers of the pair (X, D) (see Lazarsfeld [2004, p. 9.3.22]) in the interval $(0, 1]$ can be found among the roots of $b_f(s)$; see Ein, Lazarsfeld, Smith, and Varolin [2004].

For instance, it is well known that α_f can be characterized in terms of the V-filtration as

$$\alpha_f = \max \{\beta \in \mathbb{Q} \mid V^\beta \mathscr{O}_X = \mathscr{O}_X\},$$

see for instance Saito [2016, (1.2.5)], while the log canonical threshold has a similar characterization in terms of the triviality of $\mathcal{J}\big((\beta - \epsilon)D\big)$.

Assuming that f is not invertible, it is not hard to see that -1 is always a root of $b_f(s)$. The polynomial

$$\tilde{b}_f(s) = \frac{b_f(s)}{s+1}$$

is the *reduced Bernstein-Sato polynomial* of f. Inspired by (3) above and a connection with the microlocal V-filtration Saito [1994] (cf also Section 9), Saito introduced:

Definition 2.4. The *microlocal log canonical threshold* $\tilde{\alpha}_f$ is the negative of the largest root of the reduced Bernstein-Sato polynomial $\tilde{b}_f(s)$.

In particular, if $\tilde{\alpha}_f \leq 1$, then it coincides with the log canonical threshold. In other words, $\tilde{\alpha}_f$ provides a new interesting invariant precisely when the pair (X, D) is log canonical. It is already known to be related to standard types of singularities:

Theorem 2.5. *Assume that D is reduced. Then*

1. *Saito [1993, Theorem 0.4] D has rational singularities if and only if $\tilde{\alpha}_f > 1$.*
2. *Saito [2009, Theorem 0.5] D has Du Bois singularities if and only if $\tilde{\alpha}_f \geq 1$.*[1]

Example 2.6. If f is a weighted homogeneous polynomial such that x_i has weight w_i, the convention being that if f is a sum of monomials $x_1^{m_1} \cdots x_n^{m_n}$ then $\sum m_i w_i = 1$, we have $\tilde{\alpha}_f = \sum_{i=1}^n w_i$; see e.g. Saito [ibid., p. 4.1.5].

It is well known that the log canonical threshold of the pair (X, D) can be computed in terms of discrepancies; in fact, using the notation in (2.3), given any log resolution one has

$$\alpha_f = \min\{1, \gamma\}.$$

Similar precise formulas are not known for other roots of the Bernstein-Sato polynomial. Lichtin asked the following regarding the microlocal log canonical threshold.

Question 2.7. Lichtin [1989, Remark 2, p.303] Is it true that $\gamma = \tilde{\alpha}_f$?

When $\tilde{\alpha}_f \leq 1$, this is indeed the case by the discussion above. As noted by Kollár Kollár [1997, Remark 10.8], the question otherwise has a negative answer, simply due to the fact that in general the quantity on the right hand side depends on the choice of log resolution. One of the outcomes of the results surveyed in this paper will be however the inequality $\gamma \leq \tilde{\alpha}_f$; see Theorem 9.10. It would be interesting to find similar results for other roots of $\tilde{b}_f(s)$.

[1] An equivalent statement can be found in Kovács and Schwede [2011, Corollary 6.6], where it is shown that D is Du Bois if and only if the pair (X, D) is log canonical.

3 Hodge filtration on localizations

I will now start focusing on the Hodge filtration. Saito's theory of mixed Hodge modules produces useful filtered $\mathscr{D}$-modules of geometric and Hodge theoretic origin on complex varieties, which extend the notion of a variation of Hodge structure when singularities (of fibers of morphisms, of hypersurfaces, of ambient varieties, etc.) are involved; see for instance the examples in Popa [2016b, §2]. Usually the $\mathscr{D}$-module itself is quite complicated, but here we deal with one of the simplest ones.

Namely, if X is smooth complex variety of dimension n, and D is a reduced effective divisor on X, we consider the left $\mathscr{D}_X$-module

$$\mathscr{O}_X(*D) = \bigcup_{k\geq 0} \mathscr{O}_X(kD)$$

of functions with arbitrary poles along D. Locally, if $D = \mathrm{div}(f)$, then $\mathscr{O}_X(*D)$ is simply the localization $\mathscr{O}_X[f^{-1}]$, on which differential operators act by the quotient rule. This $\mathscr{D}_X$-module underlies the extension of the trivial Hodge module across D, i.e. the mixed Hodge module $j_*\mathbb{Q}_U^H[n]$, where $U = X \smallsetminus D$ and $j : U \hookrightarrow X$ is the inclusion map. A main feature of $\mathscr{D}$-modules underlying mixed Hodge modules is that they come endowed with a (Hodge) filtration, in this case $F_k\mathscr{O}_X(*D)$ with $k \geq 0$, better behaved than those on arbitrary filtered $\mathscr{D}$-modules; besides the fundamental Saito [1988], Saito [1990], see also Schnell [2014] for an introductory survey, and Sabbah and Schnell [2016] for details.

While the $\mathscr{D}$-module $\mathscr{O}_X(*D)$ is easy to understand, the Hodge filtration can be extremely complicated to describe. This is intimately linked to understanding the singularities of D and the Deligne Hodge filtration on the singular cohomology $H^\bullet(U,\mathbb{C})$. Saito [1993], Saito [2009] studied $F_k\mathscr{O}_X(*D)$ with the help of the V-filtration, and established the following results:

Theorem 3.1. *The following hold:*

1. *Saito [1993, Proposition 0.9 and Theorem 0.11] The Hodge filtration is contained in the pole order filtration, namely*

$$F_k\mathscr{O}_X(*D) \subseteq \mathscr{O}_X\big((k+1)D\big) \quad \text{for all} \quad k \geq 0,$$

and equality holds if $k \leq \tilde{\alpha}_f - 1$.

2. *Saito [2009, Theorem 0.4]* $F_0\mathscr{O}_X(*D) = \mathscr{O}_X(D) \otimes_{\mathscr{O}_X} V^1\mathscr{O}_X$.[2]

[2] This is in fact proved in *loc. cit.* with $\widetilde{V}^1\mathscr{O}_X$, the microlocal V-filtration on $\mathscr{O}_X$ (see Section 9), instead of $V^1\mathscr{O}_X$, but it can be shown that the two coincide for V^1.

Item (1) in the theorem leads to defining for each $k \geq 0$ a coherent sheaf of ideals $I_k(D)$ by the formula

$$F_k \mathscr{O}_X(*D) = \mathscr{O}_X\big((k+1)D\big) \otimes I_k(D).$$

We call these the *Hodge ideals* of the divisor D; they, and especially their extensions to $\mathbb{Q}$-divisors, play the leading role in this note.

4 Review of Hodge ideals for reduced divisors

The papers Mustaţă and Popa [2016a] and Mustaţă and Popa [2016b] are devoted to the study of Hodge ideals of reduced divisors, using both properties coming from the theory of mixed Hodge modules, and an alternative approach based on log resolutions and methods from birational geometry.

The theory is essentially complete in this case, and I will only briefly review it in this section (see also Popa [2016b, Ch. F] for a more extensive survey) and in Section 9, where the relationship with the microlocal V-filtration Saito [2016] is explained. The rest of the paper discusses the more general case of $\mathbb{Q}$-divisors, where a complete treatment is only beginning to take shape.

One may loosely summarize the main properties and results as follows:

Theorem 4.1 (Mustaţă and Popa [2016a], Mustaţă and Popa [2016b]). *Given a reduced effective divisor D on a smooth complex variety X, the sequence of Hodge ideals $I_k(D)$, with $k \geq 0$, satisfies:*
(i) $I_0(D)$ is the multiplier ideal $\mathcal{J}\big((1-\epsilon)D\big)$,[3] so in particular $I_0(D) = \mathscr{O}_X$ if and only if the pair (X, D) is log canonical. Moreover, there are inclusions

$$\cdots I_k(D) \subseteq \cdots \subseteq I_1(D) \subseteq I_0(D).$$

(ii) When D has simple normal crossings, in a neighborhood where it is given by $x_1 \cdots x_r = 0$, $I_k(D)$ is generated by $\{x_1^{a_1} \cdots x_r^{a_r} \mid 0 \leq a_i \leq k,\ \sum_i a_i = k(r-1)\}$.
(iii) D is smooth if and only if $I_k(D) = \mathscr{O}_X$ for all k; cf. also Corollary 6.5 below.
(iv) If $I_k(D) = \mathscr{O}_X$ for some $k \geq 1$ ($\Longleftrightarrow I_1(D) = \mathscr{O}_X$), then D is normal with rational singularities. More precisely, $I_1(D) \subseteq \mathrm{Adj}(D)$, the adjoint ideal of D.[4]
(v) There are non-triviality criteria for $I_k(D)$ at a point $x \in D$ in terms of the multiplicity of D at x; cf. e.g. Theorem 6.4 below.
(vi) On smooth projective varieties, $I_k(D)$ satisfy a vanishing theorem extending Nadel Vanishing for multiplier ideals (a special case of Theorem 7.1 below).

[3]Note that this follows already by combining Theorem 3.1(2) and Theorem 2.2 above.
[4]Recall that D is normal with rational singularities if and only if $\mathrm{Adj}(D) = \mathscr{O}_X$, see Lazarsfeld [2004, Proposition 9.3.48].

(vii) If H is a smooth divisor in X such that $D|_H$ is reduced, then $I_k(D)$ satisfy

$$I_k(D|_H) \subseteq I_k(D) \cdot \mathscr{O}_H,$$

with equality when H is general.
(viii) If D_1 and D_2 are reduced divisors such that $D_1 + D_2$ is also reduced, I_k satisfy the subadditivity property

$$I_k(D_1 + D_2) \subseteq I_k(D_1) \cdot I_k(D_2).$$

(ix) If $X \to T$ is a smooth family with a section $s: T \to X$, and D is a relative divisor on X such that the restriction $D_t = D|_{X_t}$ to each fiber is reduced, then

$$\{t \in T \mid I_k(D_t) \subseteq \mathfrak{m}^q_{s(t)}\}$$

is an open subset of T, for each $q \geq 1$.
(x) $I_k(D)$ determine Deligne's Hodge filtration on the singular cohomology $H^\bullet(U, \mathbb{C})$, where $U = X \smallsetminus D$, via a Hodge-to-de Rham type spectral sequence.

Note that, in view of item (i), a number of these properties are inspired by well-known properties of multiplier ideals (see Lazarsfeld [2004, Ch. 9]), though often the proofs become substantially more involved. However (ii) and (x) simply follow from standard results, via general properties of the Hodge filtration.

Another line of results proved in Mustaţă and Popa [2016a] and Mustaţă, Olano, and Popa [2017] regards the complexity of the Hodge filtration. According to Saito [2009], one says that the filtration on a $\mathscr{D}$-module $(\mathscr{M}, F_\bullet)$ is *generated at level k* if

$$F_\ell \mathscr{D}_X \cdot F_k \mathscr{M} = F_{k+\ell} \mathscr{M} \quad \text{for all} \quad \ell \geq 0.$$

The smallest integer k with this property is called the *generating level*. In the case of $\mathscr{M} = \mathscr{O}_X(*D)$ with the Hodge filtration, this can be reinterpreted as saying that

$$\mathscr{O}_X\big((k+\ell+1)D\big) \otimes I_{k+\ell}(D) = F_\ell \mathscr{D}_X \cdot \Big(\mathscr{O}_X\big((k+1)D\big) \otimes I_k(D)\Big), \tag{4.2}$$

so all higher Hodge ideals are determined by $I_k(D)$. Thus this invariant is important for concrete calculations; see e.g. Remark 9.8 below.

Theorem 4.3. *Assume that X has dimension n. Then:*

1. *Mustaţă and Popa [2016a, Theorem B] The Hodge filtration on $\mathscr{O}_X(*D)$ is generated at level $n-2$, and this bound is optimal in general.*

2. *Mustaţă, Olano, and Popa [2017, Theorem E] If D has only isolated rational singularities and $n \geq 3$, then the Hodge filtration on $\mathscr{O}_X(*D)$ is generated at level $n-3$.*

We conjecture in Mustaţă, Olano, and Popa [2017] that (2) should hold for arbitrary divisors with rational singularities. Its converse is known not to hold in general. When D has an isolated quasihomogeneous singularity, a stronger bound was given by Saito in Saito [2009, Theorem 0.7]: the generating level of $F_\bullet \mathscr{O}_X(*D)$ is $[n - \tilde{\alpha}_f - 1]$, where $\tilde{\alpha}_f$ is the microlocal log canonical threshold defined in Section 2; cf. also Theorem 2.5(1).

Example 4.4. (1) If D is a reduced divisor on a surface, then the Hodge filtration is generated at level 0, so the multiplier ideal $I_0(D)$ determines all other $I_k(D)$ via formula (4.2) for $k = 0$. See Popa [2016b, Example 13.1] for concrete calculations.

(2) If $D = (f = 0) \subset X = \mathbb{C}^3$ is a du Val surface singularity, then $I_0(D) = \mathscr{O}_X$, and since D has rational singularities, the Hodge filtration is again generated at level 0. Thus for all $k \geq 1$ we have $I_k(D) = f^{k+1} \cdot \big(F_k \mathscr{D}_X \cdot \frac{1}{f}\big)$. If however D is an elliptic singularity, then the Hodge filtration is typically not generated at level 0 any more, but only at level 1. See for instance the elliptic cone calculation in Remark 9.8.

Some first applications. The use of Hodge ideals in geometric applications is still in its early days. There are however a number of basic consequences that can already be deduced using the results above:

- Effective bounds for the degrees of hypersurfaces on which isolated singular points on a reduced hypersurface D in $\mathbb{P}^n$ of fixed degree d impose independent conditions, in the style of a classical a result of Severi for nodal surfaces in $\mathbb{P}^3$; see Mustaţă and Popa [2016a, §27]. As an example, the isolated singular points on D of multiplicity $m \geq 2$ impose independent conditions on hypersurfaces of degree $([\frac{n}{m}] + 1)d - n - 1$.
- Solution to a conjecture on the multiplicities of points on theta divisors with isolated singularities on principally polarized abelian varieties, improving in this case well-known results of Kollár and others; see Mustaţă and Popa [ibid., §29]. For instance, one shows that every point on such a theta divisor has multiplicity at most $\frac{g+1}{2}$, where g is the dimension of the abelian variety.
- Effective bound for how far the Hodge filtration coincides with the pole order filtration on the cohomology $H^\bullet(U, \mathbb{C})$ of the complement $U = X \smallsetminus D$, in the style of results of Deligne, Dimca, Saito and others. For instance, if D is a divisor with only ordinary singularities of multiplicity $m \geq 2$ in an n-dimensional X, then

$$F_p H^\bullet(U, \mathbb{C}) = P_p H^\bullet(U, \mathbb{C}) \quad \text{for all} \quad p \leq \left[\frac{n}{m}\right] - n - 1.$$

(The two filtrations on $H^\bullet(U, \mathbb{C})$ start in degree $-n$.) See Mustaţă and Popa [ibid., Theorem D].

Space constraints do not allow me to explain all of this carefully. I will however focus in detail on the second item, and in fact on an extension to pluri-theta divisors in Section 8, in order to see the machinery in action.

5 Hodge ideals for arbitrary $\mathbb{Q}$-divisors

The case of arbitrary $\mathbb{Q}$-divisors is treated in Mustaţă and Popa [2018a]. It requires a somewhat more technical setting, where the $\mathscr{D}$-modules we consider are only direct summands of $\mathscr{D}$-modules underlying mixed Hodge modules. The initial setup can be seen as a $\mathscr{D}$-module analogue of eigensheaf decompositions in the theory of cyclic covering constructions, see e.g. Esnault and Viehweg [1992, §3].

Let D be an effective $\mathbb{Q}$-divisor on X, with support Z. We denote $U = X \smallsetminus Z$ and let $j : U \hookrightarrow X$ be the inclusion map. Locally we can assume that $D = \alpha \cdot \mathrm{div}(h)$ for some nonzero $h \in \mathscr{O}_X(X)$ and $\alpha \in \mathbb{Q}_{>0}$. We denote $\beta = 1 - \alpha$.

To this data one associates by a well-known construction the left $\mathscr{D}_X$-module $\mathscr{M}(h^\beta) := \mathscr{O}_X(*Z)h^\beta$, a rank 1 free $\mathscr{O}_X(*Z)$-module with generator the symbol h^β, on which a derivation D of $\mathscr{O}_X$ acts via the rule

$$D(wh^\beta) := \left(D(w) + w\frac{\beta \cdot D(h)}{h}\right)h^\beta.$$

The case $\beta = 0$ is the localization $\mathscr{O}_X(*Z)$ considered in Section 3.

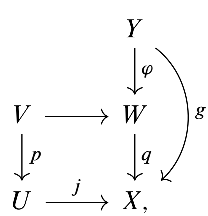
This $\mathscr{D}_X$-module does not necessarily itself underlie a Hodge module. It is however a filtered direct summand of one such, via the following construction. Let ℓ be an integer such that $\ell\beta \in \mathbb{Z}$, and consider the finite étale map

$$p\colon V = \mathbf{Spec}\, \mathscr{O}_U[y]/(y^\ell - h^{\ell\beta}) \longrightarrow U.$$

Consider also the cover

$$q\colon W = \mathbf{Spec}\, \mathscr{O}_X[z]/(z^\ell - h^{\ell\beta}) \longrightarrow X,$$

and a log-resolution $\varphi\colon Y \to W$ of the pair (W, q^*Z) that is an isomorphism over V and is equivariant with respect to the natural $\mathbb{Z}/\ell\mathbb{Z}$ action. This all fits in a commutative diagram



where the bottom square is Cartesian. Denote by E the support of $g^{-1}(Z)$.

Lemma 5.1. *Mustaţă and Popa [2018a] There is an isomorphism of filtered left $\mathscr{D}_X$-modules*

$$g_+\big(\mathscr{O}_Y(*E), F_\bullet\big) \simeq j_+ p_+(\mathscr{O}_V, F_\bullet) \simeq \bigoplus_{i=0}^{\ell-1} \big(\mathscr{M}(h^{i\beta}), F_\bullet\big),$$

where the filtration on the left hand side is given by the pushforward of the Hodge filtration in Section 3 (cf. also Theorem 4.1(iii)), while on each summand on the right hand side we consider the induced filtration.

For the notation in the lemma, recall that for any proper morphism of smooth varieties $f : X \to Y$ there is a filtered direct image functor

$$f_+ : \mathbf{D}^b\big(\mathrm{FM}(\mathscr{D}_X)\big) \longrightarrow \mathbf{D}^b\big(\mathrm{FM}(\mathscr{D}_Y)\big)$$

between the bounded derived categories of filtered $\mathscr{D}$-modules; see Saito [1990, §2.3].

Thus in this theory, the basic (local) object associated to an effective $\mathbb{Q}$-divisor D as above is the filtered $\mathscr{D}_X$-module

$$\big(\mathscr{M}(h^\beta), F_\bullet\big), \quad \text{with} \quad F_k\mathscr{M}(h^\beta) \neq 0 \Longleftrightarrow k \geq 0,$$

and for most practical purposes this has the same properties as a filtered $\mathscr{D}$-module underlying a mixed Hodge module.

One can show by a direct calculation that when the support Z is *smooth*, itself given by the equation h, then

$$F_k\mathscr{M}(h^\beta) = \mathscr{O}_X\big((k+1-[\beta])Z\big)h^\beta \subseteq \mathscr{O}_X\big((k+1)Z\big)h^\beta, \quad \text{for all} \quad k \geq 0.$$

Using this and standard reduction arguments, it follows that in general (even when Z is not necessarily defined by h), if $H = \mathrm{div}(h)$ so that $D = \alpha H$, we have

$$F_k\mathscr{M}(h^\beta) \subseteq \mathscr{O}_X\big(kZ + H\big)h^\beta, \quad \text{for all} \quad k \geq 0.$$

This allows us to formulate the following:

Definition 5.2. For each $k \geq 0$, the *k-th Hodge ideal* associated to the $\mathbb{Q}$-divisor D is defined by

$$F_k\mathscr{M}(h^\beta) = I_k(D) \otimes_{\mathscr{O}_X} \mathscr{O}_X\big(kZ + H\big)h^\beta.$$

It is standard to check that the definition of these ideals is independent of the choice of α and h, and therefore makes sense globally on X. The reduced case described in Section 3 and Section 4 corresponds to the value $\beta = 0$.

Assumption. From now on, for simplicity we assume that $\lceil D \rceil = Z$ (for instance, $D = \alpha Z$ with $0 < \alpha \leq 1$). This makes the statements more compact, while the general situation can be reduced to this case by noting that we always have

$$I_k(D) \simeq I_k(B) \otimes_{\mathscr{O}_X} \mathscr{O}_X(Z - \lceil D \rceil),$$

with $B = Z + D - \lceil D \rceil$.

In the rest of this section I will briefly explain the approach to the study of Hodge ideals based on log resolutions, originating in Mustaţă and Popa [2016a] in the reduced case, and completed in Mustaţă and Popa [2018a] in the general case. In Section 9 I will discuss the connection with the microlocal V-filtration discovered by Saito [2016], and its extension to the twisted case in Mustaţă and Popa [2018b].

Let $f : Y \to X$ be a log resolution of the pair (X, D) that is an isomorphism over $U = X \smallsetminus Z$, and denote $g = h \circ f \in \mathscr{O}_Y(Y)$. One has a filtered isomorphism

$$\big(\mathscr{M}(h^\beta), F_\bullet\big) \simeq f_+\big(\mathscr{M}(g^\beta), F_\bullet\big).$$

We use the notation $G = f^*D$ and $E = \mathrm{Supp}(G)$, the latter being a simple normal crossing divisor. It turns out that there exists a complex on Y:

$$C^\bullet_{(g^\beta, \lceil G \rceil)} : \ 0 \to \mathscr{O}_Y(-\lceil G \rceil) \otimes_{\mathscr{O}_Y} \mathscr{D}_Y \to \mathscr{O}_Y(-\lceil G \rceil) \otimes_{\mathscr{O}_Y} \Omega^1_Y(\log E) \otimes_{\mathscr{O}_Y} \mathscr{D}_Y$$

$$\to \ldots \to \mathscr{O}_Y(-\lceil G \rceil) \otimes_{\mathscr{O}_Y} \omega_Y(E) \otimes_{\mathscr{O}_Y} \mathscr{D}_Y \to 0,$$

which is placed in degrees $-n, \ldots, 0$, and such that if $x_1, \ldots, x_n$ are local coordinates, its differential is given by

$$\eta \otimes Q \to d\eta \otimes Q + \sum_{i=1}^n (dx_i \wedge \eta) \otimes \partial_i Q + (1-\beta)\big(\mathrm{dlog}(g) \wedge \eta\big) \otimes Q.$$

Moreover, this complex has a natural filtration given, for $k \geq 0$, by subcomplexes

$$F_{k-n} C^\bullet_{(g^\beta, \lceil G \rceil)} := 0 \to \mathscr{O}_Y(-\lceil G \rceil) \otimes F_{k-n}\mathscr{D}_Y \to$$

$$\to \mathscr{O}_Y(-\lceil G \rceil) \otimes \Omega^1_Y(\log E) \otimes F_{k-n+1}\mathscr{D}_Y \to \cdots \to \mathscr{O}_Y(-\lceil G \rceil) \otimes \omega_Y(E) \otimes F_k\mathscr{D}_Y \to 0.$$

The key point shown in *loc. cit.* is that the that there is a filtered quasi-isomorphism

$$(5.3) \qquad \big(C^\bullet_{g^\beta, \lceil G \rceil)}, F_\bullet\big) \simeq \big(\mathscr{M}_r(g^\beta), F_\bullet\big),$$

where

$$\mathscr{M}_r(g^\beta) := \mathscr{M}(g^\beta) \otimes_{\mathscr{O}_Y} \omega_Y \simeq h^\beta \omega_Y(*E)$$

is the filtered right $\mathscr{D}_Y$-module associated to $\mathscr{M}(g^\beta)$. In other words, the filtered complex on the left computes the Hodge filtration on $\mathscr{M}_r(g^\beta)$, hence the Hodge ideals for the simple normal crossings divisor E.

Given this fact, one can use $(C^\bullet_{(g^\beta,\lceil G\rceil)}, F_\bullet)$ as a concrete representative for computing the filtered $\mathscr{D}$-module pushforward of $\big(\mathscr{M}_r(g^\beta), F_\bullet\big)$, hence for computing the ideals $I_k(D)$. If we denote as customary by

$$\mathscr{D}_{Y\to X} = \mathscr{O}_Y \otimes_{f^{-1}\mathscr{O}_X} f^{-1}\mathscr{D}_X$$

the transfer $\mathscr{D}$-module (isomorphic to $f^*\mathscr{D}_X$ as an $\mathscr{O}_Y$-module), the result is:

Theorem 5.4. *Mustaţă and Popa [2018a] With the above notation, the following hold:*

1. *For every $p \neq 0$ and every $k \in \mathbb{Z}$, we have*

$$R^p f_*\big(C^\bullet_{(g^\beta,\lceil G\rceil)} \otimes_{\mathscr{D}_Y} \mathscr{D}_{Y\to X}\big) = 0 \quad \text{and} \quad R^p f_* F_k\big(C^\bullet_{(g^\beta,\lceil G\rceil)} \otimes_{\mathscr{D}_Y} \mathscr{D}_{Y\to X}\big) = 0.$$

2. *For every $k \in \mathbb{Z}$, the natural inclusion induces an injective map*

$$R^0 f_* F_k\big(C^\bullet_{(g^\beta,\lceil G\rceil)} \otimes_{\mathscr{D}_Y} \mathscr{D}_{Y\to X}\big) \hookrightarrow R^0 f_*\big(C^\bullet_{(g^\beta,\lceil G\rceil)} \otimes_{\mathscr{D}_Y} \mathscr{D}_{Y\to X}\big).$$

3. *We have a canonical isomorphism*

$$R^0 f_*\big(C^\bullet_{(g^\beta,\lceil G\rceil)} \otimes_{\mathscr{D}_Y} \mathscr{D}_{Y\to X}\big) \simeq \mathscr{M}_r(h^\beta)$$

that, using (2), induces for every $k \in \mathbb{Z}$ an isomorphism

$$R^0 f_* F_{k-n}\big(C^\bullet_{(g^\beta,\lceil G\rceil)} \otimes_{\mathscr{D}_Y} \mathscr{D}_{Y\to X}\big) \simeq h^\beta \omega_X\big(kZ + H\big) \otimes_{\mathscr{O}_X} I_k(D).$$

Example 5.5 ($I_0(D)$ is a multiplier ideal)**.** The lowest term in the filtration on the complex above reduces to the sheaf

$$F_{-n} C^\bullet_{(g^\beta,\lceil G\rceil)} = \omega_Y(E - \lceil f^*D\rceil)$$

in degree 0. Thus

$$I_0(D) = f_*\mathscr{O}_Y\big(K_{Y/X} + E - \lceil f^*D\rceil\big) = f_*\mathscr{O}_Y\big(K_{Y/X} - [(1-\epsilon)f^*D]\big).$$

This is by definition the multiplier ideal associated to the $\mathbb{Q}$-divisor $(1-\epsilon)D$ with $0 < \epsilon \ll 1$. Consequently (see Lazarsfeld [2004, p. 9.3.9]):

$$I_0(D) = \mathscr{O}_X \iff (X, D) \text{ is log canonical.}$$

Remark 5.6 (Local vanishing)**.** In view of Theorem 5.4(3) and Example 5.5, the statement in Theorem 5.4(1) can be seen as a generalization of Local Vanishing for multiplier ideals Lazarsfeld [ibid., Theorem 9.4.1].

Given the equivalence between the triviality of $I_0(D)$ and log canonicity, it is natural to introduce the following:

Definition 5.7. We say that the pair (X, D) is k-log canonical if

$$I_0(D) = \cdots = I_k(D) = \mathscr{O}_X.$$

Under our running assumption on D, Corollary 9.5 below implies that this is in fact equivalent to simply asking that $I_k(D) = \mathscr{O}_X$.

Example 5.8. Let Z have an ordinary singularity of multiplicity m, i.e. an isolated singular point whose projectivized tangent cone is smooth (for example the cone over a smooth hypersurface of degree m in $\mathbb{P}^{n-1}$). If $D = \alpha Z$ with $0 < \alpha \leq 1$, then (X, D) is k-log canonical if and only if $k \leq [\frac{n}{m} - \alpha]$. See Corollary 9.9, noting that $\tilde{\alpha}_f = \frac{n}{m}$ according to Saito [2009]; cf. also Mustaţă and Popa [2016a, Theorem D and Example 20.13].

Example 5.9. Irreducible theta divisors on principally polarized abelian varieties are 0-log canonical, but may sometimes not be 1-log canonical; see Mustaţă and Popa [ibid., Remark 29.3(2)]. Generic determinantal hypersurfaces are 1-log canonical, but they are not 2-log canonical; see Mustaţă and Popa [ibid., Example 20.14]. Both have rational singularities; compare with Theorem 4.1(iv).

The generation level of the Hodge filtration on $\mathscr{M}(h^\beta)$ is not well understood at the moment; for instance, depending on the value of α, examples in Mustaţă and Popa [2018a] show that on surfaces it can be either 0 or 1. It is natural to ask what is the analogue of Theorem 4.3, but also, concretely, whether the analogue of Saito's result discussed immediately after it holds:

Question 5.10. If $D = \alpha Z$, with Z reduced and having an isolated quasi-homogeneous singularity, is the generation level of the Hodge filtration on $\mathscr{M}(h^\beta)$ equal to $[n - \tilde{\alpha}_f - \alpha]$?

6 (Non)triviality criteria

The applications of the theory of multiplier ideals rely crucially on effective criteria for understanding whether they are trivial or not at a given point. The most basic are as follows; if D is an effective $\mathbb{Q}$-divisor, then:

1. If $\mathrm{mult}_x(D) \geq n = \dim X$, then $\mathcal{J}(D)_x \neq \mathscr{O}_{X,x}$; see Lazarsfeld [2004, Proposition 9.3.2].

2. If $\mathrm{mult}_x(D) < 1$, then $\mathcal{J}(D)_x = \mathcal{O}_{X,x}$; see Lazarsfeld [2004, Proposition 9.5.13].

The first is quite standard, while the second is a slightly more delicate application of inversion of adjunction.

Multiplier ideals also satisfy a birational transformation formula. If $f : Y \to X$ is any proper birational map, then

$$\mathcal{J}(D) \simeq f_*\big(\mathcal{O}_Y(K_{Y/X}) \otimes_{\mathcal{O}_Y} \mathcal{J}(f^*D)\big).$$

See Lazarsfeld [ibid., Theorem 9.2.33]. Such a compact statement is not available for higher Hodge ideals; however, using Theorem 5.4, one can show a partial analogue.

Theorem 6.1. *Mustaţă and Popa [2016a, Theorem 18.1], Mustaţă and Popa [2018a] Let $f : Y \to X$ be a projective morphism, with Y smooth. Let $Z = D_{\mathrm{red}}$, $E = (f^*D)_{\mathrm{red}}$, and denote $T_{Y/X} = \mathrm{Coker}(T_Y \to f^*T_X)$. Then:*

1. There is an inclusion

$$f_*\big(I_k(f^*D) \otimes_{\mathcal{O}_Y} \mathcal{O}_Y(K_{Y/X} + k(E - f^*Z))\big) \subseteq I_k(D).$$

2. If J is a coherent ideal on X such that $J \cdot T_{Y/X} = 0$, then

$$J^k \cdot I_k(D) \subseteq f_*\big(I_k(f^*D) \otimes_{\mathcal{O}_Y} \mathcal{O}_Y(K_{Y/X} + k(E - f^*Z))\big).$$

The first statement leads quite quickly to the following triviality criterion, in terms of the coefficients of exceptional divisors on a fixed log resolution.

Corollary 6.2. *Assume that $D = \alpha Z$ (with $0 < \alpha \le 1$) and for $f : Y \to X$ a log resolution of the pair (X, D), define γ as in (2.3). If*

$$\gamma \ge k + \alpha,$$

then $I_k(D) = \mathcal{O}_X$.

This is a key ingredient in bounding the microlocal log canonical threshold of D in terms discrepancies; see Theorem 9.10 below.

On the other hand, the second statement in Theorem 6.1 leads to nontriviality criteria that, just as in the case of multiplier ideals, are useful when combined with global statements like the vanishing theorem explained in the next section.

Corollary 6.3. *If $x \in X$ is such that $\mathrm{mult}_x Z = a$ and $\mathrm{mult}_x D = b$, and if q is a nonnegative integer such that*

$$b + ka > q + r + 2k - 1,$$

then $I_k(D) \subseteq \mathfrak{m}_x^q$. In particular, this happens if $\mathrm{mult}_W D > \frac{q+r+2k-1}{k+1}$.

At least for the moment, one can obtain somewhat stronger statements in the reduced case; the following collects some of the results in Mustaţă and Popa [2016a]. The proofs are more involved, (1) relying for instance on a deformation to ordinary singularities argument using Theorem 4.1(ix), combined with explicit calculations in that case.

Theorem 6.4. *If $x \in D$ is a point on a reduced divisor, with $m = \operatorname{mult}_x(D)$, then:*

1. *$I_k(D) \subseteq \mathfrak{m}_x^q$ if $q = \min\{m-1, (k+1)m - n\}$; see Mustaţă and Popa [ibid., Theorem E].*
2. *$I_k(D) \subseteq \mathfrak{m}_x^q$ if $m \geq 2 + \frac{q+n-2}{k+1}$; see Mustaţă and Popa [ibid., Corollary 19.4].*

As an example, for $k = 1$ the criterion in (1) can be rephrased as

$$m \geq \max\left\{q+1, \frac{n+q}{2}\right\} \implies I_1(D) \subseteq \mathfrak{m}_x^q.$$

It also implies that if $x \in D$ is a *singular* point, then

$$I_k(D) \subseteq \mathfrak{m}_x, \quad \text{for all } \ k \geq \frac{n-1}{2}.$$

In particular one obtains a smoothness criterion in terms of the Hodge filtration:

Corollary 6.5. *Mustaţă and Popa [ibid., Theorem A] The divisor D is smooth $\iff$ $I_k(D) = \mathscr{O}_X$ for all k $\iff$ $I_k(D) = \mathscr{O}_X$ for some $k \geq \frac{n-1}{2}$.*

7 Global setting and vanishing theorem

While the locally defined ideals in Definition 5.2 glue together into a global object, this is not usually the case with the $\mathscr{D}$-modules $\mathscr{M}(h^\beta)$. There is however a setting in which this can be done.

Namely, assume that $D = \frac{1}{\ell}H$, where H is an integral divisor and ℓ is a positive integer, and that there is a line bundle M such that $\mathscr{O}_X(H) \simeq M^{\otimes \ell}$. (This of course always holds when D is integral.) Let $s \in \Gamma(X, M^{\otimes \ell})$ be a section whose zero-locus is H. Recall that $U = X \smallsetminus Z$, and $j : U \hookrightarrow X$ is the inclusion. Since s does not vanish on U, we may consider the section $s^{-1} \in \Gamma(U, (M^{-1})^{\otimes \ell})$. Let

$$p \colon V = \mathbf{Spec}\big(\mathscr{O}_X \oplus M \oplus \ldots \oplus M^{\otimes(\ell-1)}\big) \longrightarrow U$$

be the étale cyclic cover corresponding to s^{-1}. The filtered $\mathscr{D}_X$-module

$$(\mathscr{M}, F_\bullet) = j_+ p_+(\mathscr{O}_V, F_\bullet)$$

underlies a mixed Hodge module, and the obvious μ_ℓ-action on $\mathscr{M}$ induces an eigenspace decomposition

$$(\mathscr{M}, F_\bullet) = \bigoplus_{i=0}^{\ell-1} (\mathscr{M}_i, F_\bullet),$$

where $\mathscr{M}_i$ is the eigenspace corresponding to the map $\lambda \to \lambda^i$, and on each $\mathscr{M}_i$ we consider the induced filtration.

On open subsets W on which M is trivialized we have isomorphisms of filtered $\mathscr{D}_W$-modules

$$\mathscr{M}_i \simeq \mathscr{M}(s_{|W}^{-i/\ell}) \quad \text{for} \quad 0 \le i \le \ell - 1,$$

which glue to a global isomorphism

$$\mathscr{M}_i \simeq M^{\otimes i} \otimes_{\mathscr{O}_X} \mathscr{O}_X(*Z) = j_* j^* M^{\otimes i}.$$

Twisting in order to globalize the $\mathscr{M}(h^\beta)$ picture, with $\beta = 1 - \frac{1}{\ell}$, we obtain global coherent ideals given by

$$F_k \mathscr{M}_i \simeq M^{\otimes i}(-H) \otimes_{\mathscr{O}_X} I_k(i/\ell \cdot H) \otimes_{\mathscr{O}_X} \mathscr{O}_X\big(kZ + H\big),$$

and the Hodge ideals $I_k(D)$ are defined by the case $i = 1$.

In this global setting, there is a vanishing theorem for Hodge ideals that in the case $k = 0$ is nothing else but the celebrated Nadel vanishing theorem for multiplier ideals. This was shown in Mustaţă and Popa [2016a, Theorem F] in the reduced case, and in Mustaţă and Popa [2018a] in general. Recall that here we are assuming $\lceil D \rceil = Z$, the support of D, for simplicity.

Theorem 7.1. *Assume that X is a smooth projective variety of dimension n, and D is a $\mathbb{Q}$-divisor as at the beginning of this section. Let L a line bundle on X such that $L + Z - D$ is ample. For some $k \ge 0$, assume that the pair (X, D) is $(k-1)$-log-canonical, i.e. $I_0(D) = \cdots = I_{k-1}(D) = \mathscr{O}_X$. Then we have:*

1. *If $k \le n$, and $L(pZ)$ is ample for all $1 \le p \le k$, then*

$$H^i\big(X, \omega_X \otimes L((k+1)Z) \otimes I_k(D)\big) = 0$$

for all $i \ge 2$. Moreover,

$$H^1\big(X, \omega_X \otimes L((k+1)Z) \otimes I_k(D)\big) = 0$$

holds if $H^j\big(X, \Omega_X^{n-j} \otimes L((k-j+1)Z)\big) = 0$ for all $1 \le j \le k$.

2. *If $k \geq n+1$ and $L((k+1)Z)$ is ample, then*

$$H^i\big(X, \omega_X \otimes L((k+1)Z) \otimes I_k(D)\big) = 0 \quad \text{for all} \quad i > 0.$$

3. *If $D + pZ$ is ample for $0 \leq p \leq k-1$, then (1) and (2) also hold with $L = M(-Z)$.*

The main ingredient in the proof is Saito's Kodaira-type vanishing theorem Saito [1990, §2.g] for mixed Hodge modules, stating that if $(\mathscr{M}, F_\bullet)$ is the filtered $\mathscr{D}$-module underlying a mixed Hodge module on a projective variety X, then

$$\mathbf{H}^i\big(X, \mathrm{gr}_k^F \mathrm{DR}(\mathscr{M}, F_\bullet) \otimes L\big) = 0 \quad \text{for all} \quad i > 0,$$

where L is any ample line bundle, and $\mathrm{gr}_k^F \mathrm{DR}(\mathscr{M}, F_\bullet)$ denotes for each k the associated graded of the induced filtration on the de Rham complex of $\mathscr{M}$. See Schnell [2016], Popa [2016a] for more on this theorem, and also Popa [2016b, §3] for a guide to interesting generalizations. In (3), this is replaced by Artin vanishing (on affine varieties) for the perverse sheaf associated to $\mathscr{M}$ via the Riemann-Hilbert correspondence.

Remark 7.2. When X has cotangent bundle with special properties, for instance when it is an abelian variety or $\mathbb{P}^n$ (or more generally a homogeneous space), the hypotheses on $(k-1)$-log canonicity and borderline Nakano-type vanishing are not needed, so vanishing holds in a completely arbitrary setting; see for instance Mustaţă and Popa [2016a, §25, §28]. Similarly, stronger vanishing holds on toric varieties Dutta [2018].

It will be important to address the following natural problem for non-reduced divisors:

Question 7.3. Does vanishing for $\mathbb{Q}$-divisors hold without the global assumption on the existence of ℓ-th roots of $\mathscr{O}_X(H)$ at the beginning of the section?

8 Example of application: pluri-theta divisors on abelian varieties

The goal here is to see explicitly how the combination of local nontriviality criteria and global vanishing for Hodge ideals can be put to use towards concrete applications. I will focus on one example: divisors in pluri-theta linear series on principally polarized abelian varieties. The result below is new, extending (and also marginally improving) part of Mustaţă and Popa [2016a, Theorem I], though the general idea is quite similar.

Let (A, Θ) be a principally polarized abelian variety of dimension g. Let $D \in |n\Theta|$ for some $n \geq 1$, whose support Z has only isolated singularities.

Theorem 8.1. *Under the hypotheses above, if $\epsilon(\Theta)$ denotes the Seshadri constant of Θ, and $x \in D$, we have:*

1. $\operatorname{mult}_x D \le n^2 g!\epsilon(\Theta) + ng!$.

2. *If A is general in the sense that $\rho(A) = 1$, or if $n = 1$, then* $\operatorname{mult}_x D \le n^2\epsilon(\Theta) + n$.

Recall that the definition of the Seshadri constant easily implies that $\epsilon(\Theta) \le \sqrt[g]{g!}$; see Lazarsfeld [2004, Proposition 5.1.9]. Before proving the theorem, let's introduce the notation $s(\ell, x)$ for the largest integer s such that the linear system $|\ell\Theta|$ separates s-jets at x, i.e. such that the restriction map

$$H^0\big(A, \mathscr{O}_A(\ell\Theta)\big) \longrightarrow H^0\big(A, \mathscr{O}_A(\ell\Theta) \otimes \mathscr{O}_A/\mathfrak{m}_x^{s+1}\big)$$

is surjective. A basic fact is that

$$\frac{s(\ell, x)}{\ell} \le \epsilon(\Theta, x), \tag{8.2}$$

the Seshadri constant of Θ at x, and that $\epsilon(\Theta, x)$ is the limit of these quotients as $\ell \to \infty$; see Lazarsfeld [ibid., Theorem 5.1.17] and its proof. Since A is homogeneous, $\epsilon(\Theta, x)$ does not actually depend on x, so it is denoted $\epsilon(\Theta)$.

Proof of Theorem 8.1. We prove (1), and at the end indicate the necessary modification needed to deduce (2). Write $D = \sum a_i Z_i$, with Z_i prime divisors, so that $Z = \sum Z_i$. Since effective divisors on abelian varieties are nef, for each i we have

$$a_i Z_i \cdot \Theta^{g-1} \le H \cdot \Theta^{g-1} = n \cdot g! \tag{8.3}$$

and since Θ is ample it follows that $a_i \le n \cdot g!$. Thus $D \le ng!Z$, so if $m = \operatorname{mult}_x(D)$, then

$$\operatorname{mult}_x(Z) \ge \lceil \frac{m}{ng!} \rceil.$$

Since x is fixed, for simplicity we denote $s_\ell = s(\ell, x)$. I claim that

$$\frac{m}{ng!} \le \lceil \frac{m}{ng!} \rceil \le \frac{(s_{n(k+1)} + g + k + 1)}{k+1}, \quad \text{for all} \quad k \ge 1. \tag{8.4}$$

Assuming the opposite inequality for some k, by Theorem 6.4(2) we have

$$I_k(Z) \subseteq \mathfrak{m}_x^{s_{n(k+1)}+2}.$$

Now according to the vanishing in Mustaţă and Popa [2016a, Theorem 28.2], a refinement on abelian varieties of the statement of Theorem 7.1 (cf. Remark 7.2), we have:

$$H^1\big(A, \mathscr{O}_A((k+1)Z) \otimes \alpha \otimes I_k(Z)\big) = 0$$

for every $\alpha \in \mathrm{Pic}^0(A)$.[5] Again using the fact that effective divisors on an abelian varieties are nef, we can write

$$(k+1)D = (k+1)Z + N,$$

where N is a nef divisor on A. On the other hand, nef line bundles on abelian varieties are special examples of what are called GV-sheaves (a condition involving the Fourier-Mukai transform, see e.g. Pareschi and Popa [2011, §2]), as they are topologically trivial twists of pullbacks of ample line bundles, so we conclude using Pareschi and Popa [ibid., Proposition 3.1][6] that we have

$$H^1\big(A, \mathscr{O}_A(n(k+1)\Theta) \otimes \alpha \otimes I_k(Z)\big) = 0 \quad \text{for all} \quad \alpha \in \mathrm{Pic}^0(A).$$

Going back to the inclusion $I_k(Z) \subseteq \mathfrak{m}_x^{s_{n(k+1)}+2}$, since Z has only isolated singularities, the quotient $\mathfrak{m}_x^{s_{n(k+1)}+2}/I_k(Z)$ is supported in dimension 0. We obtain

$$H^1\big(A, \mathscr{O}_A(n(k+1)\Theta) \otimes \alpha \otimes \mathfrak{m}_x^{s_{n(k+1)+2}}\big) = 0$$

for every $\alpha \in \mathrm{Pic}^0(A)$. But the collection of line bundles $\mathscr{O}_A(n(k+1)\Theta) \otimes \alpha$ is, as α varies in $\mathrm{Pic}^0(X)$, the same as the collection of line bundles $t_a^*\mathscr{O}_A(n(k+1)\Theta)$ as a varies in X, where t_a denotes translation by a. Therefore the vanishing above is equivalent to the statement that $|n(k+1)\Theta|$ separates $(s_{n(k+1)}+1)$-jets, which gives a contradiction and proves (8.4).

Finally, since $s_{n(k+1)} \le n(k+1)\epsilon(\Theta)$ by (8.2), we deduce that

$$m \le \frac{ng!(n(k+1)\epsilon(\Theta) + g + k + 1)}{k+1}, \quad \text{for all} \quad k \ge 1.$$

Letting $k \to \infty$, we obtain the inequality in the statement.

For the statement in (2), note that under the extra assumptions, each component Z_i satisfies $Z_i \cdot \Theta^{g-1} \ge g!$. Thus (8.3) implies the stronger bound $a_i \le n$, hence $\mathrm{mult}_x(Z) \ge \frac{m}{n}$. We can therefore eliminate the term $g!$ from all the formulas, while the rest of the argument is completely identical. □

9 V-filtration and microlocal log-canonical threshold

In this final section I turn to the connection between Hodge ideals and the V-filtration, first noted in Saito [2016]. For a $\mathbb{Q}$-divisor D on X, defined locally as $D = \alpha \cdot \mathrm{div}(f)$,

[5] Even though Z itself might not be ample, its complement $A \smallsetminus Z$ is affine, so the proof in *loc. cit.* works unchanged.

[6] The local freeness condition in the statement in *loc. cit.* is not needed in its proof.

just as in Section 5 we assume for simplicity that $Z = \mathrm{div}(f)$ is the reduced structure on D, and that $0 < \alpha \leq 1$. The corresponding statements for arbitrary D can be found in Mustaţă and Popa [2018b].

We return to the notation introduced in Section 2. Recall that for a $\mathscr{D}_X$-module $\mathscr{M}$, we denote by $\mathscr{M}_f$ its pushforward via the graph of f. In line with Saito [1993] and Saito [2016], I will use the notation

$$\mathcal{B}_f := (\mathscr{O}_X)_f, \quad \mathcal{B}_f(*Z) := (\mathscr{O}_X(*Z))_f, \quad \text{and} \quad \mathcal{B}_f^{\beta}(*Z) := (\mathscr{O}_X(*Z)f^{\beta})_f.$$

One can use the V-filtration on $\mathcal{B}_f$ in order to define some interesting ideals on X associated to D.

Definition 9.1. For each $k \geq 0$, we define

$$\tilde{I}_k(D) := \{v \in \mathscr{O}_X \mid \exists\, v_0, v_1, \ldots, v_k = v \in \mathscr{O}_X \text{ such that } \sum_{i=0}^{k} v_i \otimes \partial_t^i \in V^{\alpha}\mathcal{B}_f\} \subseteq \mathscr{O}_X.$$

Since $0 < \alpha \leq 1$, this is just another way of writing the filtration $\widetilde{V}^{\bullet}\mathscr{O}_X$ induced on $\mathscr{O}_X$ by Saito's *microlocal V-filtration* Saito [1994], Saito [2016]. In the notation of *loc. cit.*, we have

$$\tilde{I}_k(D) = \widetilde{V}^{k+\alpha}\mathscr{O}_X. \tag{9.2}$$

When $D = Z$ is a reduced divisor (i.e. $\alpha = 1$), a comparison theorem between Hodge ideals and these "microlocal" ideals was established recently by Saito.

Theorem 9.3. *Saito [ibid., Theorem 1] If D is reduced, then for every $k \geq 0$ we have*

$$I_k(D) = \tilde{I}_k(D) \mod f.$$

The statement means that the equality happens only in the quotient $\mathscr{O}_D$. For $k = 0$ it holds without modding out by f, by Theorem 2.2. However, for higher k it does not necessarily hold in $\mathscr{O}_X$; see Remark 9.8.

Its extension to arbitrary $\mathbb{Q}$-divisors is established in Mustaţă and Popa [2018b], as a consequence of a statement which is more explicit, in the sense of completely computing Hodge ideals in terms of the V-filtration, even in the reduced case. For $i \geq 0$, we denote

$$Q_i(X) = \prod_{j=0}^{i-1}(X + j) \in \mathbb{Z}[X].$$

Theorem 9.4. *Mustaţă and Popa [ibid.] If D is a $\mathbb{Q}$-divisor as above, then for every $k \geq 0$ we have*

$$I_k(D) = \left\{ \sum_{j=0}^{p} Q_j(\alpha) f^{p-j} v_j \mid \sum_{j=0}^{p} v_j \otimes \partial_t^j \delta \in V^\alpha \mathcal{B}_f \right\}.$$

In particular, we have

$$I_k(D) = \tilde{I}_k(D) \mod f.$$

One of the key technical points in Mustaţă and Popa [ibid.] is a description of the V-filtration on $\mathcal{B}_f^\beta(*Z)$ in terms of that on $\mathcal{B}_f(*Z)$, based on Sabbah's computation of the V-filtration in terms of the Bernstein-Sato polynomials of individual elements in the $\mathscr{D}$-module Sabbah [1987].

Theorem 9.4 has consequences regarding the basic behavior of Hodge ideals that, surprisingly, at the moment are not known by other means. Recall for instance the chain of inclusions in Theorem 4.1(i); this seems unlikely to hold in the general $\mathbb{Q}$-divisor case, but the following is nevertheless true, given (9.2).

Corollary 9.5. *For each $k \geq 1$ we have*

$$I_k(D) + (f) \subseteq I_{k-1}(D) + (f).$$

Stronger statements hold for the first nontrivial ideal, as it is not hard to see that the k-log-canonicity of a divisor D (see Definition 5.7) implies that $(f) \subseteq I_{k+1}(D)$.

Corollary 9.6. *If (X, D) is $(p-1)$-log canonical, then*

$$\tilde{I}_p(D) \subseteq I_p(D) = \tilde{I}_p(D) + (f)$$

and also

$$I_{p+1}(D) \subseteq I_p(D).$$

In particular, we always have $I_1(D) \subseteq I_0(D)$.

Another important consequence regards the behavior of the Hodge ideals $I_k(\alpha Z)$ when α varies. In the case of I_0, it is well known that they get smaller as α increases, and that there is a discrete set of values of α (called jumping coefficients) where there the ideal actually changes; see Lazarsfeld [2004, Lemma 9.3.21]. This is not the case for higher k; for the cusp $Z = (x^2 + y^3 = 0)$ and $5/6 < \alpha \leq 1$, one can see that

$$I_2(\alpha Z) = (x^3, x^2y^2, xy^3, y^4 - (2\alpha + 1)x^2 y),$$

and thus we obtain incomparable ideals. However, Theorem 9.4 implies that the picture does becomes similar to that for multiplier ideals if one considers the images in $\mathscr{O}_D$.

Corollary 9.7. *Given any k, there exists a finite set of rational numbers $0 = c_0 < c_1 < \cdots < c_s < c_{s+1} = 1$ such that for each $0 \le i \le s$ and each $\alpha \in (c_i, c_{i+1}]$ we have*

$$I_k(\alpha Z) \bmod f = I_k(c_{i+1}Z) \bmod f = \text{constant}$$

and such that

$$I_k(c_{i+1}Z) \bmod f \subseteq I_k(c_i Z) \bmod f.$$

In fact, for a fixed k, the set of c_i is contained in the set of jumping coefficients for the V-filtration on $\mathcal{B}_f$ in the interval $(k, k+1]$.

Remark 9.8 (Calculations)**.** There are also significant computational consequences; indeed, in Saito [2016, §2.2-2.4], Saito fully computes the microlocal V-filtration for weighted-homogeneous isolated singularities. For example, in the case of diagonal hypersurfaces $f = x_1^{a_1} + \cdots + x_n^{a_n}$ (which was previously obtained in Maxim, Saito, and Schuermann [2016, Example 2.6] using a Thom-Sebastiani type theorem), $\widetilde{V}^\alpha$ is generated by monomials of the form $x_1^{\nu_1} \cdots x_n^{\nu_n}$ satisfying

$$\sum_{i=1}^{n} \frac{1}{a_i}\left(\nu_i + 1 + \left[\frac{\nu_i}{a_i - 1}\right]\right) \ge \alpha.$$

Saito also shows in *loc. cit.* that $I_1(D) = \tilde{I}_1(D)$ in the reduced homogeneous case, though this typically fails for $k \ge 2$. Consider as an example the elliptic cone $D = (x^3 + y^3 + z^3 = 0) \subseteq \mathbf{A}^3$. The pair is log canonical, hence $I_0(D) = \mathcal{O}_X$. Moreover, it follows from the above that

$$I_1(D) = \tilde{I}_1(D) = (x^2, y^2, z^2, xyz).$$

Theorem 4.3(1) implies that from this one can compute all other $I_k(D)$. The calculations in Saito [2016] show however that the element $-2x^4 + xy^3 + xz^3$ belongs to $I_2(D)$, but not to $\tilde{I}_2(D)$. Many concrete calculations of Hodge ideals can also be performed based on the results in Saito [2009]; see also the upcoming Zhang [2018] for generalizations to $\mathbb{Q}$-divisors.

Microlocal log canonical threshold. Part of the usefulness of the results above stems from the connection between the (microlocal) V-filtration and the Bernstein-Sato polynomial of f and its roots; cf. Section 9. Most importantly for us here, and by analogy with the description of the log canonical threshold in terms of $V^\bullet \mathcal{O}_X$, one has

$$\tilde{\alpha}_f = \max\{\gamma \in \mathbb{Q} \mid \widetilde{V}^\gamma \mathcal{O}_X = \mathcal{O}_X\},$$

see for instance Saito [2016, (1.3.8)]. Therefore Theorem 9.4 immediately implies the following formula for the log canonicity index of D, obtained first in Saito [ibid.] when $\alpha = 1$; recall that we are assuming $D = \alpha Z$ with $0 < \alpha \le 1$.

Corollary 9.9. *Let*

$$p_0 := \min\{p \mid I_p(D) \neq \mathscr{O}_X\} = \max\{p \mid (X, D) \text{ is } (p-1)\text{–log canonical}\}.$$

Then $p_0 = [\tilde{\alpha}_f - \alpha + 1]$.

Corollary 9.9 can be combined with the information in Corollary 6.2, coming from the birational description of Hodge ideals, in order to obtain the inequality

$$\gamma < [\tilde{\alpha}_f - \alpha] + \alpha + 1, \quad \text{for all } \ 0 < \alpha \leq 1.$$

Going back to Question 2.7 and the subsequent comments, optimizing as α varies we obtain the following partial positive answer to Lichtin's question:

Theorem 9.10. *Mustaţă and Popa [2018b] We have* $\gamma \leq \tilde{\alpha}_f$.

One can also use the methods explained here in order to give bounds on $\tilde{\alpha}_f$, the first of which is a result of Saito [1994]:

Corollary 9.11. *Let* $n = \dim X$. *Then:*

1. $\tilde{\alpha}_f \leq [n/2]$.
2. *If* Z *has only singular points of multiplicity at most* m, *whose projectivized tangent cone satisfies* $\dim \operatorname{Sing}\big(\mathbb{P}(C_x D)\big) \leq r$, *then* $[\tilde{\alpha}_f] \geq \frac{n-r-1}{m}$.

Indeed, if this weren't the case in (1), then by Corollary 9.9 for $\alpha = 1$ we would have $I_k(Z) = \mathscr{O}_X$ for some $k \geq \frac{n-1}{2}$. But then Corollary 6.5 implies that Z is smooth, a contradiction. For (2) one uses the bound for $I_k(Z) = \mathscr{O}_X$ in Mustaţă and Popa [2016b, Corollary D], based on Theorem 4.1(vii).

This bound in (1) is optimal, since by Example 2.6 for a quadric $f = x_1^2 + \cdots + x_n^2$ one has $\tilde{\alpha}_f = n/2$. Saito shows in fact in Saito [1994, Theorem 0.4] that all the negatives of the roots of $\tilde{b}_f(s)$ are contained in the interval $[\tilde{\alpha}_f, n - \tilde{\alpha}_f]$. Alternatively, one can see this result, together with Theorem 9.3, as recovering Corollary 6.5.

References

Nero Budur and Morihiko Saito (2005). "Multiplier ideals, V-filtration, and spectrum". *J. Algebraic Geom.* 14.2, pp. 269–282. MR: 2123230 (cit. on pp. 800, 802).

Jean-Pierre Demailly (1993). "A numerical criterion for very ample line bundles". *J. Differential Geom.* 37.2, pp. 323–374. MR: 1205448 (cit. on p. 799).

Alexandru Dimca, Philippe Maisonobe, and Morihiko Saito (2011). "Spectrum and multiplier ideals of arbitrary subvarieties". *Ann. Inst. Fourier (Grenoble)* 61.4, 1633–1653 (2012). MR: 2951747 (cit. on p. 800).

Yajnaseni Dutta (2018). *Vanishing for Hodge ideals on toric varieties*. In preparation (cit. on p. 817).

Lawrence Ein, Robert Lazarsfeld, and Michael Nakamaye (1996). "Zero-estimates, intersection theory, and a theorem of Demailly". In: *Higher-dimensional complex varieties (Trento, 1994)*. de Gruyter, Berlin, pp. 183–207. MR: 1463179 (cit. on p. 799).

Lawrence Ein, Robert Lazarsfeld, Karen E. Smith, and Dror Varolin (2004). "Jumping coefficients of multiplier ideals". *Duke Math. J.* 123.3, pp. 469–506. MR: 2068967 (cit. on p. 803).

Hélène Esnault and Eckart Viehweg (1992). *Lectures on vanishing theorems*. Vol. 20. DMV Seminar. Birkhäuser Verlag, Basel, pp. vi+164. MR: 1193913 (cit. on p. 809).

Michel Granger (2010). "Bernstein-Sato polynomials and functional equations". In: *Algebraic approach to differential equations*. World Sci. Publ., Hackensack, NJ, pp. 225–291. MR: 2766095 (cit. on p. 803).

Masaki Kashiwara (1983). "Vanishing cycle sheaves and holonomic systems of differential equations". In: *Algebraic geometry (Tokyo/Kyoto, 1982)*. Vol. 1016. Lecture Notes in Math. Springer, Berlin, pp. 134–142. MR: 726425 (cit. on pp. 801, 802).

– (1976/77). "B-functions and holonomic systems. Rationality of roots of B-functions". *Invent. Math.* 38.1, pp. 33–53. MR: 0430304 (cit. on pp. 800, 803).

János Kollár (1997). "Singularities of pairs". In: *Algebraic geometry—Santa Cruz 1995*. Vol. 62. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 221–287. MR: 1492525 (cit. on pp. 800, 803, 804).

Sándor J. Kovács and Karl E. Schwede (2011). "Hodge theory meets the minimal model program: a survey of log canonical and Du Bois singularities". In: *Topology of stratified spaces*. Vol. 58. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, pp. 51–94. MR: 2796408 (cit. on p. 804).

Robert Lazarsfeld (2004). *Positivity in algebraic geometry. II*. Vol. 49. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Positivity for vector bundles, and multiplier ideals. Springer-Verlag, Berlin, pp. xviii+385. MR: 2095472 (cit. on pp. 802, 803, 806, 807, 812–814, 818, 821).

Ben Lichtin (1989). "Poles of $|f(z,w)|^{2s}$ and roots of the b-function". *Ark. Mat.* 27.2, pp. 283–304. MR: 1022282 (cit. on pp. 800, 803, 804).

Bernard Malgrange (1983). "Polynômes de Bernstein-Sato et cohomologie évanescente". In: *Analysis and topology on singular spaces, II, III (Luminy, 1981)*. Vol. 101. Astérisque. Soc. Math. France, Paris, pp. 243–267. MR: 737934 (cit. on pp. 801, 802).

Laurenţiu Maxim, Morihiko Saito, and Joerg Schuermann (Oct. 2016). "Thom–Sebastiani theorems for filtered D-modules and for multiplier ideals". arXiv: 1610.07295 (cit. on p. 822).

Mircea Mustaţă, Sebastian Olano, and Mihnea Popa (2017). "Local vanishing and Hodge filtration for rational singularities". To appear in J. Inst. Math. Jussieu. arXiv: 1703.06704 (cit. on pp. 807, 808).

Mircea Mustaţă and Mihnea Popa (May 2016a). "Hodge ideals". To appear in Mem. Amer. Math. Soc. arXiv: 1605.08088 (cit. on pp. 800, 806–808, 811, 813–818).

– (June 2016b). "Restriction, subadditivity, and semicontinuity theorems for Hodge ideals". To appear in Int. Math. Res. Not. arXiv: 1606.05659 (cit. on pp. 800, 806, 823).

– (2018a). *Hodge ideals for $\mathbb{Q}$-divisors I: birational approach*. In preparation (cit. on pp. 800, 809–814, 816).

– (2018b). *Hodge ideals for $\mathbb{Q}$-divisors II: V-filtration*. In preparation (cit. on pp. 800, 811, 820, 821, 823).

Giuseppe Pareschi and Mihnea Popa (2011). "Regularity on abelian varieties III: relationship with generic vanishing and applications". In: *Grassmannians, moduli spaces and vector bundles*. Vol. 14. Clay Math. Proc. Amer. Math. Soc., Providence, RI, pp. 141–167. MR: 2807853 (cit. on p. 819).

Giuseppe Pareschi, Mihnea Popa, and Christian Schnell (2017). "Hodge modules on complex tori and generic vanishing for compact Kähler manifolds". *Geom. Topol.* 21.4, pp. 2419–2460. MR: 3654112 (cit. on p. 800).

Mihnea Popa (2016a). "Kodaira–Saito vanishing and applications". *Enseign. Math.* 62.1-2, pp. 49–89. arXiv: 1407.3294. MR: 3605809 (cit. on p. 817).

– (May 2016b). "Positivity for Hodge modules and geometric applications". In: *Proceedings of the 2015 AMS Summer Institute, Salt Lake City*, p. 31. arXiv: 1605.08093 (cit. on pp. 800, 805, 806, 808, 817).

Mihnea Popa and Christian Schnell (2013). "Generic vanishing theory via mixed Hodge modules". *Forum Math. Sigma* 1, e1, 60. MR: 3090229 (cit. on p. 800).

– (2014). "Kodaira dimension and zeros of holomorphic one-forms". *Ann. of Math. (2)* 179.3, pp. 1109–1120. MR: 3171760 (cit. on p. 800).

– (2017). "Viehweg's hyperbolicity conjecture for families with maximal variation". *Invent. Math.* 208.3, pp. 677–713. MR: 3648973 (cit. on p. 800).

Claude Sabbah (1987). "D-modules et cycles évanescents (d'après B. Malgrange et M. Kashiwara)". In: *Géométrie algébrique et applications, III (La Rábida, 1984)*. Vol. 24. Travaux en Cours. Hermann, Paris, pp. 53–98. MR: 907935 (cit. on pp. 801, 803, 821).

Claude Sabbah and Christian Schnell (2016). *The MHM project* (cit. on p. 805).

Morihiko Saito (1988). "Modules de Hodge polarisables". *Publ. Res. Inst. Math. Sci.* 24.6, 849–995 (1989). MR: 1000123 (cit. on pp. 799, 801, 802, 805).

Morihiko Saito (1990). "Mixed Hodge modules". *Publ. Res. Inst. Math. Sci.* 26.2, pp. 221–333. MR: 1047415 (cit. on pp. 799, 805, 810, 817).
– (1993). "On b-function, spectrum and rational singularity". *Math. Ann.* 295.1, pp. 51–74. MR: 1198841 (cit. on pp. 804, 805, 820).
– (1994). "On microlocal b-function". *Bull. Soc. Math. France* 122.2, pp. 163–184. MR: 1273899 (cit. on pp. 801, 803, 804, 820, 823).
– (2009). "On the Hodge filtration of Hodge modules". *Mosc. Math. J.* 9.1, 161–191, back matter. MR: 2567401 (cit. on pp. 804, 805, 807, 808, 813, 822).
– (Dec. 2016). "Hodge ideals and microlocal V-filtration". arXiv: 1612.08667 (cit. on pp. 800, 803, 806, 811, 819, 820, 822).
Christian Schnell (2012). "Complex analytic Néron models for arbitrary families of intermediate Jacobians". *Invent. Math.* 188.1, pp. 1–81. MR: 2897692 (cit. on p. 800).
– (May 2014). "An overview of Morihiko Saito's theory of mixed Hodge modules". arXiv: 1405.3096 (cit. on p. 805).
– (2016). "On Saito's vanishing theorem". *Math. Res. Lett.* 23.2, pp. 499–527. MR: 3512896 (cit. on p. 817).
Yum-Tong Siu (2004). "Hyperbolicity in complex geometry". In: *The legacy of Niels Henrik Abel*. Springer, Berlin, pp. 543–566. MR: 2077584 (cit. on p. 799).
Botong Wang (2016). "Torsion points on the cohomology jump loci of compact Kähler manifolds". *Math. Res. Lett.* 23.2, pp. 545–563. MR: 3512898 (cit. on p. 800).
Chuanhao Wei (Nov. 2017). "Logarithmic Kodaira dimension and zeros of holomorphic log-one-forms". arXiv: 1711.05854 (cit. on p. 800).
Tamaki Yano (1983). "b-functions and exponents of hypersurface isolated singularities". In: *Singularities, Part 2 (Arcata, Calif., 1981)*. Vol. 40. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 641–652. MR: 713287 (cit. on pp. 800, 803).
Mingyi Zhang (2018). In preparation (cit. on p. 822).

Received 2017-12-01.

Mihnea Popa
Department of Mathematics
Northwestern University
2033 Sheridan Road
Evanston, IL 60208
USA
mpopa@math.northwestern.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (825–848)

INTERACTION BETWEEN SINGULARITY THEORY AND THE MINIMAL MODEL PROGRAM

Chenyang Xu (许晨阳)

Abstract

We survey some recent topics on singularities, with a focus on their connection to the minimal model program. This includes the construction and properties of dual complexes, the proof of the ACC conjecture for log canonical thresholds and the recent progress on the 'local stability theory' of an arbitrary Kawamata log terminal singularity.

1 Introduction

Through out this paper, we will consider algebraic singularities in characteristic 0. It is well known that even if we are mostly interested in smooth varieties, for many different reasons, we have to deal with singular varieties. For the minimal model program (MMP) (also known as Mori's program), the reason is straightforward, mildly singular varieties are built into the MMP process, and there is no good way to avoid them (see e.g. Kollár and Mori [1998]). In fact, the development of the MMP has been intertwined with the progress of our understanding of the corresponding singularity theory, in particular for the classes of singularities preserved by an MMP sequence. This is one of the main reasons why when the theory was started around four decades ago, people spent a lot of time to classify these singularities. However, once we move beyond dimension three, an explicit characterisation of these singularities is often too complicated, and we have to search for a more intrinsic and qualitative method. It turns out that MMP theory itself provides many new tools for the study of singularities. In this note, we will survey some recent progress along these lines. More precisely, we will discuss the construction and properties of dual complexes, the proof of the ACC conjecture for log canonical thresholds, and the recently developed concept of 'local stability theory' of an arbitrary Kawamata log

The author is partially supported by 'The National Science Fund for Distinguished Young Scholars (11425101).
MSC2010: primary 14J17; secondary 14E30, 14J45.

terminal singularity. We hope these different aspects will give the reader an insight to the modern philosophy of studying singularities from the MMP viewpoint.

In the rest of the introduction, we will give a very short account on some of the main ideas. Given a singularity $x \in X$ in characteristic 0, the first birational model that one probably thinks of is a smooth one given by Hironaka's theorem on resolution of singularities. However, started from dimension three, there are often too many possible resolutions and examples clearly suggest that in a general case, an 'optimal resolution' does not exist. By the philosophy of MMP, we should run a sequence of relative MMP, which allows us to start from a general birational model over $x \in X$, e.g., an arbitrary resolution, and produce a sequence of relative birational models. The output of this MMP is a birational model, which usually is mildly singular but still equipped with many desirable properties. Furthermore, since during the MMP process, each step is a simple surgery like a divisorial contraction or a flip, we can keep track of many properties of the models and use this information to answer questions. As an example, in Section 2, we will consider the construction of a CW-complex as a topological invariant for an isolated singularity $x \in X$ with K_X being $\mathbb{Q}$-Cartier, namely the dual complex of a minimal resolution denoted by $\mathfrak{DMR}(x \in X)$.

A possibly more profound principle is that there is a local-to-global analogue between different types of singularities and the building blocks of varieties. More precisely, the MMP can be considered as a process to transform and decompose an arbitrary projective variety into three types, which respectively have positive (Fano), zero (Calabi-Yau) or negative (KSBA) first Chern class. These three classes are naturally viewed as building blocks for higher dimensional varieties. As a local counterpart, we consider normal singularities whose canonical class is $\mathbb{Q}$-Cartier. There is a closely related trichotomy: the minimal log discrepancy is larger, equal or smaller than 0. In fact, guided by the local to global principle, we are able to discover striking new results on singularities. In Section 3, we will focus on the proof of Shokurov's ACC conjecture on log canonical thresholds, which is achieved via an intensive interplay between local and global geometry. In Section 4, we will investigate in a new perspective on Kawamata log terminal (klt) singularities which are precisely the singularities with positive log discrepancies and form the local analog of Fano varieties. We will explain some deep insights on klt singularities inspired by advances in the study of Fano varieties. More precisely, for Fano varieties, we have the notion of K-(semi,poly)stability which has a differential geometry origin, as it is expected to characterise the existence of a Kähler-Einstein metric. For klt singularities, the local to global principle leads us to discover a (conjectural) stability theory, packaged in the *Stable Degeneration Conjecture* 4.4, which can be considered as a local analogue to the K-stability for Fano varieties.

Reference: Giving a comprehensive account of the relation between the singularity theory and the MMP is far beyond the scope of this note. The singularity theory in the MMP is extensively discussed in the book Kollár [2013b]. Ever since Kollár's book was published, many different aspects of singularity theory have significantly evolved, and important new results have been established.

Acknowledgments. I want to thank my coauthors, especially Christopher Hacon, János Kollár, Chi Li and James McKernan for many discussions and joint works on the mathematical materials surveyed in this note. I am grateful to Christopher Hacon and János Kollár for long lists of valuable comments.

2 Dual complex

There are many standard references in the subject of MMP, see e.g. Kollár and Mori [1998]. Here we recall some basic definitions. Given a normal variety X and a $\mathbb{Q}$-divisor Δ whose coefficients along prime components are contained in $\mathbb{Q}\cap[0,1]$, we call (X,Δ) a *$\mathbb{Q}$-Cartier log pair* if $K_X+\Delta$ is $\mathbb{Q}$-Cartier, e.g. there is some positive integer N such that $N(K_X+\Delta)$ is Cartier. For a divisorial valuation E whose centre on X is non-empty, we can assume there is a birational model $f: Y\to X$, such that E is a divisor on Y. Then we can define the *discrepancy* $a(E,X,\Delta)$ for a $\mathbb{Q}$-Cartier log pair (X,Δ) to be the multiplicity of

$$K_{Y/X}+f^*\Delta=K_Y-f^*(K_X+\Delta)$$

along E. This is a rational number of the form $\frac{p}{N}$ for some integer p. For many questions, it is more natural to look at the *log discrepancy* $A(E,X,\Delta)=a(E,X,\Delta)+1$, which is also denoted by $A_{X,\Delta}(E)$ in the literature. We say that (X,Δ) is *log canonical* (resp. *Kawamata log terminal* (*klt*)) if $A(E,X,\Delta)\geq 0$ (resp. $A(E,X,\Delta)>0$) for all divisorial valuations E whose centre $\mathrm{Center}_X(E)$ on X is non-empty. There is another important class called *divisorial log terminal* (*dlt*) sitting in between: a log pair (X,Δ) is dlt if there is a smooth open locus $U\subset X$, such that $\Delta_U=_{\mathrm{defn}}\Delta|_U$ is a reduced divisor satisfying (U,Δ_U) is simple normal crossing, and any divisor E with the centre $\mathrm{Center}_X(E)\subset X\setminus U$ satisfies $A(E,X,\Delta)>0$. The main property for the discrepancy function is that $a(E,X,\Delta)$ monotonically increases under a MMP sequence, which implies that the MMP will preserve the classes of singularities defined above (cf. Kollár and Mori [ibid., pp. 3.42–3.44]).

We call a projective variety X to be *a $\mathbb{Q}$-Fano variety* if X only has klt singularities and $-K_X$ is ample. Similarly, a projective pair (X,Δ) is called *a log Fano pair* if (X,Δ) is klt and $-K_X-\Delta$ is ample.

2.1 Dual complex as PL-homeomorphism invariant. For a simple normal crossing variety E, it is natural to consider how the components intersect with each other. This combinatorial data is captured by the dual complex $\mathcal{D}(E)$ (see Definition 2.1). A typical example one can keep in mind is the dual graph $\mathcal{D}(E)$ for a resolution $(Y, E) \to X$ of a normal surface singularity, where the exceptional curve E is assumed to be of simple normal crossings. This invariant is indispensable for the study of surface singularities (see e.g. Mumford [1961]). Nevertheless, the concept of dual complex can be defined in a more general context.

Definition 2.1 (Dual Complex)**.** Let $E = \bigcup_{i \in I} E_i$ be a pure dimensional scheme with irreducible components E_i. Assume that

1. each E_i is normal and

2. for every $J \subset I$, if $\cap_{i \in J} E_i$ is nonempty, then every connected component of $\cap_{i \in J} E_i$ is irreducible and has codimension $|J| - 1$ in E.

Note that assumption (2) implies the following.

3. For every $j \in J$, every irreducible component of $\cap_{i \in J} E_i$ is contained in a unique irreducible component of $\cap_{i \in J \setminus \{j\}} E_i$.

The *dual complex* $\mathcal{D}(E)$ of E is the regular cell complex obtained as follows. The vertices are the irreducible components of E and to each irreducible component of $W \subset \cap_{i \in J} E_i$ we associate a cell of dimension $|J| - 1$. This cell is usually denoted by v_W. The attaching map is given by condition (3). Note that $\mathcal{D}(E)$ is a simplicial complex iff $\cap_{i \in J} E_i$ is irreducible (or empty) for every $J \subset I$.

Fixed a dlt pair (X, Δ), the reduced part $E =_{\text{defn}} \Delta^{=1}$ of Δ satisfies the assumptions in Definition 2.1 (see e.g. Kollár [2013b, Section 4.2]), thus we can define $\mathcal{D}(X, \Delta) =_{\text{defn}} \mathcal{D}(E)$. Clearly, by the definition of U in the definition of dlt singularity (X, Δ), we can pick any such U, then $\mathcal{D}(X, \Delta) = \mathcal{D}(\Delta|_U)$. Furthermore, if two dlt pairs (X, Δ) and (X', Δ') are crepant birationally equivalent, i.e., the pull backs of $K_X + \Delta$ and $K_{X'} + \Delta'$ to a common model are the same, then applying the weak factorisation theorem to log resolutions of (X, Δ) and (X', Δ') and carefully tracking the dual complex given by divisors with log discrepancy 0 on each birational model, we can show that $\mathcal{D}(X, \Delta)$ and $\mathcal{D}(X', \Delta')$ are PL-homeomorphic (see de Fernex, Kollár, and Xu [2017, p. 11]).

Given a sequence of MMP

$$(X_1, \Delta_1) \dashrightarrow (X_2, \Delta_2) \dashrightarrow \cdots \dashrightarrow (X_k, \Delta_k),$$

as the log discrepancies of (X_i, Δ_i) monotonically increase, we have

$$\mathcal{D}(X_1, \Delta_1) \supset \mathcal{D}(X_2, \Delta_2) \supset \cdots \supset \mathcal{D}(X_k, \Delta_k).$$

Remark 2.2. Although part of the MMP, including the abundance conjecture, remains to be conjectural, all the MMP results we need in this note are already proved in Birkar, Cascini, Hacon, and McKernan [2010] and its extensions, e.g. Hacon and Xu [2013].

We know the following technical but useful criterion.

Lemma 2.3 (de Fernex, Kollár, and Xu [2017, p. 19]). *If for a step $(X_i, \Delta_i) \dashrightarrow (X_{i+1}, \Delta_{i+1})$ of an MMP sequence, the extremal ray R_i satisfies that $R_i \cdot D_i > 0$ for a component D_i of $\Delta_i^{=1}$, then $\mathfrak{D}(X_i, \Delta_i) \supset \mathfrak{D}(X_{i+1}, \Delta_{i+1})$ is a homotopy equivalence.*

Now we can apply this to various geometric situations. We first consider the application to the study of a singularity $x \in X \subset \mathbb{C}^N$. It has been known for long time (see Milnor [1968]) that all local topological information of $x \in X$ is encoded in the link defined as

$$\text{Link}(x \in X) =_{\text{defn}} X \cap B_\epsilon(x)$$

for a sufficiently small radius ϵ. Following the strategy of studying surfaces (as in e.g. Mumford [1961]), we pick a log resolution $Y \to (x \in X)$ and let $E =_{\text{defn}} f^{-1}(x))$ (in particular, E is simple normal crossing). Then $\text{Link}(x \in X)$ is a tubular neighbourhood of E and $\mathfrak{D}(E)$ contains some key information of this tubular structure.

Example 2.4. Consider the well known classification of rational double points (or Du Val singularities) on surface:

1. Type A_n: $x^2 + y^2 + z^{n+1} = 0$.
2. Type D_n: $x^2 + zy^2 + z^{n-1} = 0$ $(n \geq 4)$.
3. Type E_6: $x^2 + y^3 + z^4 = 0$.
4. Type E_7: $x^2 + y(y^2 + z^3) = 0$
5. Type E_8: $x^2 + y^3 + z^5 = 0$.

Then the minimal resolution Y with the exceptional locus E forms a log resolution, and $\mathfrak{D}(E)$ is the graph underlying the corresponding Dynkin diagram.

Using the weak factorisation theorem Abramovich, Karu, Matsuki, and Włodarczyk [2002], one shows that the homotopy class of $\mathfrak{D}(E)$ is a well-defined homotopy invariant $\mathfrak{DR}(x \in X)$ which does not depend on the choice of the log resolution (Y, E) (see e.g. Payne [2013]). The strategy in our previous discussion then can be used to show the following result.

Theorem 2.5. *For an isolated normal singularity $x \in X$ with K_X being $\mathbb{Q}$-Cartier, we can define a canonical PL-homeomorphism invariant $\mathfrak{DMR}(x \in X)$ which has the homotopy class of $\mathfrak{DR}(x \in X)$.*

Proof. First we take a log resolution $(Y, E) \to X$ which is isomorphic outside $X \setminus \{x\}$, then run a relative MMP of (Y, E) over X. The output $(X^{\rm dlt}, \Delta^{\rm dlt})$ is called *a dlt modification* of $x \in X$. Then we define a regular complex

$$\mathfrak{DMR}(x \in X) =_{\rm defn} \mathfrak{D}(\Delta^{\rm dlt}).$$

Since different dlt modifications are crepant birationally equivalent to each other, we know $\mathfrak{DMR}(x \in X)$ gives a well defined PL-homeomorphism class by the discussion before. Furthermore, we can deduce from Lemma 2.3 that for all the birational models appearing in steps of the relative MMP, including the last one $(X^{\rm dlt}, \Delta^{\rm dlt})$, the dual complexes have the same homotopy type. Then it implies $\mathfrak{DMR}(x \in X)$ is homotopy equivalent to $\mathfrak{DR}(x \in X)$. □

The regular complex $\mathfrak{DMR}(x \in X)$ can be considered as a geometric realisation of the weight 0 part of the Hodge theoretic invariant attached to $x \in X$. An interesting corollary to Theorem 2.5 is that if we consider a klt singularity $x \in X$, then $\mathfrak{DR}(x \in X)$ is contractible, as in the special case of Example 2.4.

Another natural context in which the dual complex appears is for the motivic zeta function using the log resolution formula (cf. Denef and Loeser [2001, Section 3]). The techniques developed here can be used to show that the only possible maximal order pole of the motivic zeta function is the negative of the log canonical threshold, which was a conjecture by Veys (see Nicaise and Xu [2016a]).

2.2 Dual complex of log Calabi-Yau pairs. Similar ideas can be applied when we consider the setting of a proper degeneration $Y \to C$ of projective varieties over a smooth pointed curve $(C, 0)$. Here we consider the dual complex $\mathfrak{D}(Y_0^{\rm red})$ where $Y_0^{\rm red}$ is the reduced fiber over 0 and assume $(Y, Y_0^{\rm red})$ is a dlt pair.

For subjects like mirror symmetry, the degeneration of Calabi-Yau varieties is of particular interests. From a birational geometry view, if we consider a family $\pi : Y \to C$, and assume a general fiber has $K_{Y_t} \sim_{\mathbb{Q}} 0$, then after running an MMP over C (cf. Fujino [2011]), we end up with a model Y which satisfies $K_Y + Y_0^{\rm red} \sim_{\mathbb{Q}} 0$. Then for such models with this extra condition, any two of them are crepant birationally equivalent which implies $\mathfrak{D}(Y_0^{\rm red})$ is well defined up to PL-homeomorphism.

Indeed in this case, the topological invariant $\mathfrak{D}(Y_0^{\rm red})$ is first defined by Kontsevich-Soibelman as the 'essential skeleton' of the Berkovich theoretic non-archimedean analytification $Y^{\rm an}$ (see Kontsevich and Soibelman [2001], Mustaţă and Nicaise [2015], and Nicaise and Xu [2016b]), and it plays an important role in the study of the algebro-geometric version of the SYZ conjecture (cf. Strominger, Yau, and Zaslow [1996], Kontsevich and Soibelman [2001], and Gross and Siebert [2011] etc.). The same argument for proving Theorem 2.5 can be used to show the essential skeleton $\mathcal{D}(Y_0^{\mathrm{red}})$ is homotopy equivalent to Y^{an} (see Nicaise and Xu [2016b]).

To understand $\mathcal{D}(Y_0^{\mathrm{red}})$, we first need to describe its local structure, i.e., for a prime component $X \subset Y_0$, to understand the link of the corresponding vertex v_X in $\mathcal{D}(Y_0^{\mathrm{red}})$. It is given by $\mathcal{D}(\Delta^{=1})$ where Δ is defined by formula

$$(K_Y + Y_0^{\mathrm{red}})|_X = K_X + \Delta \sim_{\mathbb{Q}} 0.$$

Our goal is to show under suitable conditions, the dual complex coming from a (log) Calabi-Yau variety is close to simple objects like a sphere or a disc. In fact, we can show the following.

Theorem 2.6 (Kollár and Xu [2016])**.** *Let* (X, Δ) *be a projective dlt pair which satisfies* $K_X + \Delta \sim_{\mathbb{Q}} 0$. *Assume* $\dim(\mathcal{D}(\Delta^{=1})) > 1$, *then the following holds.*

1. $H^i(\mathcal{D}(\Delta^{=1}), \mathbb{Q}) = 0$ *for* $1 \le i \le \dim(\mathcal{D}(\Delta^{=1}, \mathbb{Q}))$.
2. $\mathcal{D}(\Delta^{=1})$ *is a pseudo-manifold with boundary (Kollár and Kovács [2010]).*
3. *There is a natural surjection* $\pi_1(X^{\mathrm{sm}}) \twoheadrightarrow \pi_1(\mathcal{D}(\Delta^{=1}))$.
4. *The profinite completion* $\hat{\pi}_1(\mathcal{D}(\Delta^{=1}))$ *is finite.*

Proof. We sketch the proof under the extra assumption that $\dim(\mathcal{D}(\Delta^{=1}))$ is maximal, i.e. equal to $n-1$. Then (1) is easily obtained using Hodge theory. We need to apply MMP theory to show (2) and (3). Here we explain the argument for (3). A carefully chosen MMP process (see Kollár and Xu [2016, Section 6]) allows us to change the model from X to a birational model X', with the property that if we define the effective $\mathbb{Q}$-divisor Δ' on X' to be the one such that (X, Δ) and (X', Δ') are crepant birationally equivalent, then the support of Δ' contains an ample divisor. From this we conclude that

$$\pi_1(X'^{\mathrm{sm}}) \twoheadrightarrow \pi_1(\Delta'^{=1}) \twoheadrightarrow \pi_1(\mathcal{D}(\Delta^{=1})),$$

where the first surjection follows from the (singular version of) Lefschetz Hyperplane Theorem. We also have $\pi_1(X^{\mathrm{sm}}) \to \pi_1(X'^{\mathrm{sm}})$ by tracking the MMP process, which concludes (3). Finally, (4) follows from Xu [2014]. (We note here that we indeed expect $\pi_1(\mathcal{D}(\Delta^{=1}))$ is finite, but to apply the above argument, we need $\pi_1(X^{\mathrm{sm}})$ for the underlying variety X of a log Fano pair. For now, we only know its pro-finite completion is finite.) □

A remaining challenging question is to understand the torsion cohomological group $H^i(\mathcal{D}(\Delta^{=1}), \mathbb{Z})$ for a dlt log Calabi-Yau pair (X, Δ).

3 ACC of log canonical thresholds

Given a holomorphic function f with $f(0)=0$, the complex singular index

$$c(f)=\sup\{c \mid \frac{1}{|f|^c} \text{ is locally } L^2\text{-integrable at } 0\}$$

introduced by Arnold is a fundamental invariant, which appears in many contexts (see Arnol'd, Guseĭn-Zade, and Varchenko [1985, II. Chap. 13]). In a more general setting, in birational geometry, this invariant is interpreted as the *log canonical threshold* of an effective $\mathbb{Q}$-divisor D with respect to a log pair (X,Δ)

$$\mathrm{lct}(X,\Delta;D)=\max\{t \mid (X,\Delta+tD) \text{ is log canonical}\}.$$

Using a log resolution, it is not hard to show that when X is the local germ $0\in\mathbb{C}^n$, $\Delta=0$ and $D=(f)$, $\mathrm{lct}(X;D)=c(f)$.

Example 3.1. Let $X=\mathbb{C}^n$, $f=x_1^{m_1}+\cdots+x_n^{m_n}$, then by Kollár [1997, p. 8.15]

$$\mathrm{lct}(\mathbb{C}^n,f)=\min\{1,\sum_{i=1}^{n}\frac{1}{m_i}\}.$$

For a fixed n, all such numbers form an infinite set which satisfies the ascending chain condition.

See Kollár [ibid., Section 8-10] for a wonderful survey, including relations with other branches of mathematics.

We define the following set.

Definition 3.2. Fix the dimension n and two sets of positive numbers I and J, we denote by $\mathrm{LCT}_n(I,J)$ the set consisting of all numbers $\mathrm{lct}(X,\Delta;D)$ such that $\dim(X)=n$, the coefficients of Δ are in I and the coefficients of D are in J.

Our main contribution to the study of log canonical thresholds is showing the following theorem.

Theorem 3.3 (Hacon, McKernan, and Xu [2014, Theorem 1.1], ACC Conjecture for log canonical thresholds)**.** *If I and J satisfy the descending chain condition (DCC), then* $\mathrm{LCT}_n(I,J)$ *satisfies the ascending chain condition (ACC).*

In such a generality, this was conjectured in Shokurov [1992], although in a lot of earlier works, questions of a similar flavour already appeared. For $X=\mathbb{C}^n$ (or even more generally for bounded singularities) and $D=(f)$, this was solved by de Fernex, Ein, and Mustaţă [2010] using a different approach. In fact, while our proof is via global geometry, the argument in de Fernex, Ein, and Mustaţă [ibid.] uses a more local method.

To understand our strategy, we start with a well-known construction: Given a log canonical pair (X, Δ) with a prime divisor E over X with the log discrepancy $A(E, X, \Delta) = 0$, if X admits a boundary Δ' such that (X, Δ') is klt, then applying the MMP we can construct a model $f : Y \to X$ such that $\mathrm{Ex}(f)$ is equal to the divisor E (see Birkar, Cascini, Hacon, and McKernan [2010, p. 1.4.3]). Denote by $\Delta_Y = E + f_*^{-1}\Delta$ and restrict $K_Y + \Delta_Y$ to a general fiber F of $f : E \to f(E)$. Since E has coefficient one in Δ_Y, the adjunction formula says there is a boundary Δ_F such that

$$K_F + \Delta_F = (K_Y + \Delta_Y)|_F = f^*(K_X + \Delta) \sim_{\mathbb{Q}} 0.$$

In other words, using the model Y constructed by an MMP technique, from a lc pair (X, Δ) which is not klt along a subvariety $f(E)$, we obtain a log Calabi-Yau pair (F, Δ_F) of smaller dimensional.

We note that even in the case $\Delta_Y = E$, since Y could be singular along codimension 2 points on E, it is not always the case that $\Delta_E = 0$. Nevertheless, if the coefficients of Δ are in a set $I \subset [0, 1]$, then the coefficients of Δ_F are always in the set

$$D(I) =_{\mathrm{defn}} \{\frac{n-1+a}{n} \mid n \in \mathbb{N}, a = \sum_{i=1}^{j} a_i \text{ where } a_i \in I\} \cap [0, 1]$$

(see e.g. Kollár [2013b, p. 3.45]). In particular, if I satisfies the DCC, then $D(I)$ satisfies the DCC. This is why we work with such a general setting of coefficients as it works better with the induction.

Moreover, if there is a sequence of pairs (X_i, Δ_i) and strictly increasing log canonical thresholds t_i with respect to the divisors D_i, then the above construction will produce a sequence of log Calabi-Yau varieties (F_i, Δ_{F_i}) corresponding to $(X_i, \Delta_i + t_i D_i)$ with the property that the restriction of $f_*^{-1}(\Delta_i + t_i D_i)$ on F_i yields components of Δ_{F_i} with strictly increasing coefficients as $i \to \infty$. Therefore, to get a contradiction, it suffices to prove the following global version of the ACC conjecture.

Theorem 3.4 (Hacon, McKernan, and Xu [2014, Theorem 1.5])**.** *Fix n and a DCC set I, then there exists a finite set $I_0 \subset I$ such that for any projective n-dimensional log canonical Calabi-Yau pair (X, Δ), i.e. $K_X + \Delta \sim_{\mathbb{Q}} 0$, with the coefficients of Δ contained in I, it indeeds holds that the coefficients of Δ are in I_0.*

We note that Theorem 3.4 in dimension $n-1$ implies Theorem 3.3 in dimension n. More crucially, Theorem 3.4 changes the problem from a local setting to a global one and we have many new tools to study it. In particular, as we will explain below, Theorem 3.4

relates to the boundedness results on log general type pairs. This is a central topic in the study of such pairs, especially for the construction of the compact moduli space of KSBA stable pairs, which is the higher dimensional analogue of the moduli space of marked stable curves $\overline{\mathfrak{M}}_{g,n}$ (see e.g. Kollár [2013a] and Hacon, McKernan, and Xu [2016]).

Since I satisfies the DCC, if such a finite set I_0 does not exist, we can construct an infinite sequences (X_i, Δ_i) of log canonical Calabi-Yau pairs of dimension at most n, such that after reordering, if we write $\Delta_i = \sum_{j=1}^{k} a_i^j \Delta_i^j$, $\{a_i^j\}_{i=1}^{\infty}$ monotonically increases for any fixed $1 \le j \le k$ and strictly increases for at least one. Furthermore, after running an MMP, we can reduce to the case that the underlying variety X_i is a Fano variety with the Picard number $\rho(X_i) = 1$. Then if we push up the coefficients of Δ_i to get a new boundary

$$\Delta_i' =_{\text{defn}} \sum_{j=1}^{k} a_\infty^j \Delta_i^j \qquad \text{where } a_\infty^j = \lim_i a_i^j,$$

$K_{X_i} + \Delta_i'$ is ample. By enlarging I, we can start with the assumption that all accumulation points of I are also contained I. In particular, the coefficients $a_\infty^j \in I$. Moreover, recall that by induction on the dimension, we can assume Theorem 3.4 holds for dimension $n-1$, which implies Theorem 3.3 in dimension n. Thus for i sufficiently large, (X_i, Δ_i') is also log canonical. Then we immediately get a contradiction to the second part of (2) in the following theorem.

Theorem 3.5 (Hacon, McKernan, and Xu [2014, Theorem 1.3])**.** *Fix dimension n and a DCC set $I \subset [0,1]$. Let $\mathfrak{D}_n(I)$ be the set of all pairs*

$$\{(X, \Delta) \mid \dim(X) = n, (X, \Delta) \textit{ is lc and the coefficients of } \Delta \textit{ are in } I\},$$

and $\mathfrak{D}_n^\circ(I) \subset \mathfrak{D}_n(I)$ the subset of pairs with $K_X + \Delta$ being big. Then the following holds.

1. *The set* $\mathrm{Vol}_n(I) = \{\mathrm{vol}(K_X + \Delta) \mid (X, \Delta) \in \mathfrak{D}_n(I)\}$ *satisfies DCC.*
2. *There exists a positive integer $N = N(n, I)$ depending on n and I such that the linear system $|N(K_X + \Delta)|$ induces a birational map for any $(X, \Delta) \in \mathfrak{D}_n^\circ(I)$. Moreover, there exists $\delta > 0$ depending only on n and I, such that if $(X, \Delta) \in \mathfrak{D}_n^\circ(I)$, then $K_X + (1-\delta)\Delta$ is big.*

The part (1) was a conjecture of Alexeev-Kollár (cf. Kollár [1994] and Alexeev [1994]). As already mentioned, it is the key in the proof of the boundedness of the moduli space of KSBA stable pairs with fixed numerical invariants. See Hacon, McKernan, and Xu [2016] for a survey on this topic and related literature.

During the proof of Theorem 3.5, we have to treat (1) and (2) simultaneously. Such a strategy was first initiated in Tsuji [2007], and carried out by Hacon and McKernan [2006] and Takayama [2006] for X with canonical singularities and $\Delta = 0$. It started with the simple observation that for smooth varieties of general type, birationally boundedness implies boundedness (after the MMP is settled). In Hacon, McKernan, and Xu [2013], we prove a log version of this, which says that log birational boundedness essentially implies Theorem 3.5. This is significantly harder, and we use ideas from Alexeev [1994] which established the two dimensional case of Theorem 3.5. After this, it remains to show that all pairs in $\mathfrak{D}^{\circ}_n(I)$ with the volume bounded from above by an arbitrarily fixed constant is always log birationally bounded, which is done in Hacon, McKernan, and Xu [2014]. One key ingredient is to produce appropriate boundaries on the log canonical centres such that the classical techniques of inductively cutting log canonical centres initiated in Angehrn and Siu [1995] can be followed here.

Addressing the proof for the ACC of the log canonical thresholds in this circle of global questions is a crucial idea in our solution to it. In fact, in the pioneering work McKernan and Prokhorov [2004], an attempt was already made to establish a connection between the ACC of log canonical thresholds and a global question on boundedness but for the set of K_X-negative varieties, i.e. Fano varieties. More precisely, it has been shown in McKernan and Prokhorov [ibid.] that the ACC conjecture of log canonical thresholds is implied by Borisov-Alexeev-Borisov (BAB) conjecture which is about the bounededness of Fano varieties with a uniform positive lower bound on log discrepancies. More recently, the BAB conjecture is proved in Birkar [2016].

In Hacon, McKernan, and Xu [2014], under a suitable condition on I and assuming that $J = \{\mathbb{N}\}$, we show that the accumulation points of $\mathrm{LCT}_n(I) =_{\mathrm{defn}} \mathrm{LCT}_n(I, \mathbb{N})$ are contained in $\mathrm{LCT}_{n-1}(I)$, confirming the Accumulation Conjecture due to Kollár.

It attracts considerable interests to find out the effective bound for the constants appearing in Theorem 3.3 and Theorem 3.5. So far it is only successful for low dimension. For instance when $I = 0$, Theorem 3.3 implies that there exists an optimal $\delta_n < 1$ such that $\mathrm{LCT}_n =_{\mathrm{defn}} \mathrm{LCT}(\{0\}) \subset [0, \delta_n] \cup \{1\}$, and $(\delta_n, 1)$ is called *the n-dimensional gap*. It is known $\delta_2 = \frac{5}{6}$, but δ_3 is unknown. In Kollár [1997, p. 8.16], it is asked whether

$$\delta_n = 1 - \frac{1}{a_n} \qquad \text{where } a_1 = 2, a_i = a_1 \cdots a_{i-1} + 1.$$

Our approach in general only gives the existence of δ_n.

Remark 3.6 (ACC Conjecture on minimal log discrepancy)**.** There is another deep conjecture about ACC properties of singularities due to Shokurov, which seems to be still far open.

ACC Conjecture of mld: Given a log canonical singularity $x \in (X, \Delta)$, we can define

$$\mathrm{mld}_{X,\Delta}(x) = \min\{A_{X,\Delta}(E) \mid \mathrm{Center}_E(X) = \{x\}\}.$$

If we fix a finite set I, and it is conjectured that the set

$$\mathrm{MLD}_n(I) = \{\mathrm{mld}_{X,\Delta}(x) | \ \dim(X) = n, \text{ coefficients of } \Delta \text{ are in I}\}$$

satisfies the ACC.

However, compared to the log canonical thresholds, what kind of global questions connect to this conjecture still remains to be in a myth. For instance, it is not clear which special geometric structure is carried by a divisor attaining the minimal log discrepancy.

4 Klt singularities and K-stability

In this section, our discussion will focus on klt singularities. When klt singularities were first introduced, they appeared to be just a technical tool to prove results in the MMP. However, it has become more and more clear that the klt singularities form a very interesting class of singularities, which naturally appears in many context besides the MMP such as constructing Kähler-Einstein metrics of Fano varieties etc..

In particular, philosophically, it has been clear that there is an analogy between klt singularities and Fano varieties. Traditionally, people often prove some properties for an arbitrary Fano variety, then figure out what they imply for the cone singularity over a Fano variety, and finally generalise the statements to any klt singularity. Only after the corresponding MMP results are established (e.g. Birkar, Cascini, Hacon, and McKernan [2010]), such analogy can be carried out in a more concrete manner by really attaching suitable global objects, e.g. Fano varieties, to the singularities. The first construction was the plt blow up (cf. e.g. Xu [2014]) which for a given klt singularity $x \in (X, \Delta)$, one constructs a birational model $f : Y \to X$ such that f is isomorphic outside x, $f^{-1}(x)$ is an irreducible divisor S, and $(Y, S + f_*^{-1}\Delta)$ is plt. We can also assume $-S$ is ample over X, and then (S, Δ_S) is a log Fano pair, where

$$K_S + \Delta_S =_{\mathrm{defn}} (K_Y + S + f_*^{-1}\Delta)|_S.$$

The divisor S in this construction is called *a Kollár component*. It was used to show some local topological properties of $x \in X$ including $\mathfrak{DR}(x \in X)$ is contractible (de Fernex, Kollár, and Xu [2017]), and the pro-finite completion $\hat{\pi}_1^{\mathrm{loc}}(x \in X)$ of the local fundamental group

$$\pi_1^{\mathrm{loc}}(x \in X) =_{\mathrm{defn}} \pi_1(\mathrm{Link}(x \in X))$$

is finite (Xu [2014] and Z. Tian and Xu [2016]). However, given a klt singularity, usually there could be many Kollár components over it. Only until the circle of ideas of local stability were introduced in Li [2015b], a more canonical picture, though some parts still remain conjectural, becomes clear. In what follows we give a survey on this topic.

Definition 4.1 (Valuations)**.** Let R be an n-dimensional regular local domain essentially of finite type over a ground field k of characteristic zero. Then a (real) valuation v of $K = \mathrm{Frac}(R)$ is any map $v\colon K^* \to \mathbb{R}$ which satisfies the following properties for all a, b in K^*:

1. $v(ab) = v(a) + v(b)$,

2. $v(a+b) \geq \min(v(a), v(b))$, with equality if $v(a) \neq v(b)$.

Let $(X, x) = (\mathrm{Spec}(R), \mathfrak{m})$, we denote the space of valuations

$$\mathrm{Val}_{X,x} = \{\text{real valuations } v \text{ of } K \text{ with } v(f) > 0 \text{ for any } f \in \mathfrak{m}\}.$$

It has a natural topology (see Jonsson and Mustaţă [2012, Section 4.1]).

If (X, Δ) is klt, following Jonsson and Mustaţă [ibid., Section 5], we can define the function of log discrepancy $A_{X,\Delta}(v)$ on $\mathrm{Val}_{X,x}$ extending the log discrepancy of divisorial valuations defined in Section 2, and we denote by $\mathrm{Val}^{=1}_{X,x} \subset \mathrm{Val}_{X,x}$ the subset consisting of all valuations with log discrepancy equal to 1. Similar to the global definition of volumes, we can also define a local volume of a valuation for $v \in \mathrm{Val}_{X,x}$ (see Ein, Lazarsfeld, and Smith [2003])

$$\mathrm{vol}(v) = \lim \frac{\mathrm{length}(R/\mathfrak{a}_k)}{k^n/n!},$$

where $\mathfrak{a}_k = \{f \in R \mid v(f) \geq k\}$.

Definition 4.2 (Li [2015b])**.** For any valuation $v \in \mathrm{Val}_{X,x}$, we define the *normalised volume* $\widehat{\mathrm{vol}}_{X,\Delta}(v) = (A_{X,\Delta}(v))^n \cdot \mathrm{vol}(v)$, and the volume of the klt singularity $x \in (X, \Delta)$ to be $\mathrm{vol}(x, X, \Delta) = \inf_{v \in \mathrm{Val}_{X,x}} \widehat{\mathrm{vol}}(v)$. By abuse of notation, we will often denote $\mathrm{vol}(x, X, \Delta)$ by $\mathrm{vol}(x, X)$ if the context is clear.

It is easy to see that $\widehat{\mathrm{vol}}(v) = \widehat{\mathrm{vol}}(\lambda v)$ for any $\lambda > 0$, so that we can only consider the function $\widehat{\mathrm{vol}}$ on $\mathrm{Val}^{=1}_{X,x}$. In Li [ibid.], it was shown that $\mathrm{vol}(x, X) > 0$. In Liu [2016], a different characterisation is given:

$$\mathrm{vol}(x, X) = \inf_{\mathfrak{m}-\text{primary } \mathfrak{a}} \mathrm{mult}(\mathfrak{a}) \cdot \mathrm{lct}(X, \Delta; \mathfrak{a})^n. \tag{1}$$

See Lazarsfeld [2004, p. 9.3.14] for the definition of the log canonical threshold of a klt pair (X, Δ) with respect to an ideal $\mathfrak{a}$. Then in Blum [2016], using an argument combining estimates on asymptotic invariants and the generic limiting construction, it is show that there always exists a valuation v such that $\mathrm{vol}(x, X) = \widehat{\mathrm{vol}}(v)$, i.e., the infimum is indeed a minimum, confirming a conjecture in Li [2015b]. Therefore the main questions left are two-fold.

Question 4.3. For a klt singularity $x \in (X, \Delta)$,

I. Characterise the geometric properties of the minimiser v.

II. Compute the volume $\mathrm{vol}(x, X)$.

In what follows below, we will discuss these two questions in different sections.

4.1 Geometry of the minimiser. In the recent birational geometry study of Fano varieties, it has become clear that the interplay between the ideas from higher dimensional geometry and the ideas from the complex geometry, centred around the study of Kähler-Einstein metrics, will lead to deep results. The common ground is the notion of K-(semi, poly)stability and their cousin definitions (see e.g. Odaka [2013], Li and Xu [2014], and Fujita [2015] etc.). An example is the construction of a proper moduli scheme parametrising the smoothable K-polystable Fano varieties (see e.g. Li, Wang, and Xu [2014]). Although to establish a moduli space of Fano varieties is certainly a natural question to algebraic geometers, without a condition like K-stability with a differential geometry origin, such a functor does not behave well (e.g. the functor of smooth family Fano manifolds is not seperated.). Moreover the arguments used in the current construction of moduli spaces of K-polystable Fano varieties heavily depend on the results proved using analytic tools as in Chen, Donaldson, and Sun [2015] and G. Tian [2015].

Our main motivation to consider v is to establish a 'local K-stability' theory for klt singularities, guided by the local-to-global philosophy mentioned in the introduction. In particular, we propose the following conjecture for all klt singularities.

Conjecture 4.4 (Stable Degeneration Conjecture, Li [2015b] and Li and Xu [2017]). *Given any arbitrary klt singularity $x \in (X = \mathrm{Spec}(R), \Delta)$. There is a unique minimiser v up to rescaling. Furthermore, v is quasi-monomial, with a finitely generated associated graded ring $R_0 =_{\mathrm{defn}} \mathrm{gr}_v(R)$, and the induced degeneration*

$$(X_0 = \mathrm{Spec}(R_0), \Delta_0, \xi_v)$$

is a K-semistable Fano cone singularity. (See below for the definitions.)

For the definition of quasi-monomial valuations, see Jonsson and Mustaţă [2012, Section 3]. It is shown that they are the same as Abhyankar valuations (Ein, Lazarsfeld, and Smith [2003, p. 2.8]). From an arbitrary quasi-monomial valuation $v \in \mathrm{Val}_{X,x}$, there is a standard process to degenerate $\mathrm{Spec}(R)$ to the associated graded ring $\mathrm{Spec}(R_0)$ over a complicated (e.g. non-Noetherian) base (see Teissier [2003]). However, when R_0 is finitely generated, the degeneration can be understood in a much simpler way: we can embed $\mathrm{Spec}(R)$ into an affine space $\mathbb{C}^N$ of sufficiently large dimension, such that there exists a $\mathbb{C}^*$-action on $\mathbb{C}^N$ with a suitable weight $(\lambda_1, ..., \lambda_N)$ satisfying that $\mathrm{Spec}(R_0)$ is the degeneration of $\mathrm{Spec}(R)$ under this one-parameter $\mathbb{C}^*$-action (see e.g. Li and Xu [2017]).

The following example which predates our study is a prototype from the context of constructing Sasaki-Einstein metrics in Sasaki geometry.

Example 4.5 (Fano cone singularity)**.** Assume that $X = \mathrm{Spec}_{\mathbb{C}}(R)$ is a normal affine variety. Denote by T a complex torus $(\mathbb{C}^*)^r$ which acts on X faithfully. Let $N = \mathrm{Hom}(\mathbb{C}^*, T) \cong \mathbb{Z}^{\oplus r}$ be the co-weight lattice and $M = N^*$ the weight lattice. We have a weight space decomposition

$$R = \bigoplus_{\alpha \in \Gamma} R_\alpha \text{ where } \Gamma = \{\alpha \in M |\ R_\alpha \neq 0\}.$$

We assume $R_{(0)} = \mathbb{C}$ which means there is a unique fixed point o contained in the closure of each orbit. Denote by $\sigma^\vee \subset M_{\mathbb{R}}$ the convex cone generated by Γ, which is called the *weight cone* (or the *moment cone*). We define the *Reeb cone*

$$\mathfrak{t}^+_{\mathbb{R}} := \{\ \xi \in N_{\mathbb{R}}\ \ |\ \langle \alpha, \xi \rangle > 0 \text{ for any } \alpha \in \Gamma\}.$$

Then for any vector $\xi \in \mathfrak{t}^+_{\mathbb{R}}$ on X we can associate a natural valuation v_ξ, which is given by

$$v_\xi(f) = \min\{\langle \alpha, \xi \rangle \mid f_\alpha \neq 0 \text{ if we write } f = \sum f_\alpha\}.$$

If X have klt singularities, we call (X, ξ) a *Fano cone singularity* for the following reason: for any $\xi \in N_{\mathbb{Q}} \cap \mathfrak{t}^+_{\mathbb{R}}$, then it generates a $\mathbb{C}^*$-action on X, and the quotient will be a log Fano pair as we assume X is klt.

For isolated Fano cone singularities, minimising the normalised volume $\widehat{\mathrm{vol}}$ among all valuations of the form v_ξ ($\xi \in \mathfrak{t}^+_{\mathbb{R}}$) was initiated in the work Martelli, Sparks, and Yau [2008], where $\widehat{\mathrm{vol}}$ is defined analytically. It is shown there that the existence of a Sasaki-Einstein metric along ξ_0 implies v_{ξ_0} is a minimiser among all $\xi \in \mathfrak{t}^+_{\mathbb{R}}$. Moreover, it is proved that $\widehat{\mathrm{vol}}$ is strictly convex on $\mathfrak{t}^+_{\mathbb{R}}$, which is an evidence for the claim of the uniqueness in Conjecture 4.4.

Later, in Collins and Székelyhidi [2015], following G. Tian [1997] and Donaldson [2001], the K-(semi)polystability was formulated for Fano cone singularities. It was a straightforward calculation from the definition to show that if (X, ξ_0) is K-semistable, then v_{ξ_0} is a minimiser among all valuations of the form v_ξ for $\xi \in \mathfrak{t}^+_{\mathbb{R}}$. However, it takes significant more work in Li and Xu [2017] to show that if (X, ξ_0) is K-semistable, then v_{ξ_0} is a minimiser in the much larger space $\mathrm{Val}_{X,x}$ and unique among all quasi-monomial valuations up to rescaling (see Step 3 and 6 in the sketch of the proofs of Theorem 4.6 and Theorem 4.7 below).

For a normal singularity $x \in (X = \mathrm{Spec}(R), \Delta)$ with a quasi-monomial valuation $v \in \mathrm{Val}_{X,x}$ of rational rank r, we assume that its associated graded ring R_0 is finitely generated. By the grading, $\mathrm{Spec}(R_0)$ admits a torus $T \cong (\mathbb{C}^*)^r$-action, thus we can put it in (a log generalisation of) the setting of Example 4.5 as follows. Let Φ be the valuative semigroup of v, then it generates a group $\Phi^g \cong \mathbb{Z}^r$ which is isomorphic to the weight lattice $M = N^*$. Under this isomorphism the weight cone is generated by $\alpha \in \Phi$. Since the embedding $\iota_v \colon \Phi^g \to \mathbb{R}$ restricts to $\iota_v^+ \colon \Phi \to \mathbb{R}_+$, it yields a vector in the Reeb cone $\mathfrak{t}^+_{\mathbb{R}} \subset N_{\mathbb{R}}$, denoted by ξ_v. Let Δ_0 be the natural divisorial degeneration of Δ on $X_0 = \mathrm{Spec}(R_0)$. We call such a valuation $v \in \mathrm{Val}_{x,X}$ to be *K-semistable*, if (X_0, Δ_0, ξ_v) is a K-semistable Fano cone. In particular, we require (X_0, Δ_0) to be klt. Since a K-semistable valuation is always a minimiser (see Theorem 4.7), Conjecture 4.4 predicts that for any klt singularity $x \in (X, \Delta)$, the minimiser of $\widehat{\mathrm{vol}}$ is precisely the same as the notion of a K-semistable valuation.

We have established various parts of Conjecture 4.4. First we consider the case that the minimiser is a divisorial valuation.

Theorem 4.6 (Li and Xu [2016, Theorem 1.2])**.** *Let $x \in (X, \Delta)$ be a klt singularity. If a divisorial valuation $\mathrm{ord}_S \in \mathrm{Val}_{X,x}$ minimises the function $\widehat{\mathrm{vol}}_{X,\Delta}$, then S is a Kollár component over x, and the induced log Fano pair (S, Δ_S) is K-semistable. Furthermore, $\widehat{\mathrm{vol}}(\mathrm{ord}_S) < \widehat{\mathrm{vol}}(\mathrm{ord}_E)$ for any divisor $E \neq S$ centred on x.*

Conversely, if S is a Kollár component centred on x such that the induced log Fano pair (S, Δ_S) is K-semistable, then ord_S minimises $\widehat{\mathrm{vol}}_{X,\Delta}$.

An immediate consequence is that, if instead of searching general Kollár components, we only look for the semi-stable ones, then if one exists, it is unique. In general, Conjecture 4.4 predicts that if we choose a sequence of rational vectors $v_i \in \mathfrak{t}^+_{\mathbb{Q}}$ that converge to v, then the quotient of X_0 by the $\mathbb{C}^*$-action along v_i induces a Kollár component S_i centred on $x \in (X, \Delta)$ which satisfies $c_i \cdot \mathrm{ord}_{S_i} \to \xi_v$ after a suitable rescaling (cf. Li [2015a] and Li and Xu [2017]). In Li and Xu [2016], we confirm that any minimiser is always a limit of a sequence of Kollár components with a suitable rescaling.

In general, a quasi-monomial valuation with higher rational rank could appear as the minimiser (cf. Blum [2016]). In this case, we can also prove the following result.

Theorem 4.7 (Li and Xu [2017, Theorem 1.1])**.** *Let $x \in (X, \Delta)$ be a klt singularity. Let v be a quasi-monomial valuation in $\mathrm{Val}_{X,x}$ that minimises $\widehat{\mathrm{vol}}_{(X,\Delta)}$ and has a finitely generated associated graded ring $\mathrm{gr}_v(R)$. Then the following properties hold:*

(a) The degeneration $\big(X_0 =_{\mathrm{defn}} \mathrm{Spec}\big(\mathrm{gr}_v(R)\big), \Delta_0, \xi_v\big)$ is a K-semistable Fano cone, i.e. v is a K-semistable valuation;

(b) Let v' be another quasi-monomial valuation in $\mathrm{Val}_{X,x}$ that minimises $\widehat{\mathrm{vol}}_{(X,\Delta)}$. Then v' is a rescaling of v.

Conversely, any quasi-monomial valuation that satisfies (a) above is a minimiser.

Sketch of ideas in the Proofs of Theorem 4.6 and Theorem 4.7. The proof consists of a few steps, involving different techniques.

Step 1: In this step, we illustrate how Kollár components come into the picture. From each ideal $\mathfrak{a}$, we can take a dlt modification of

$$f : (Y, \Delta_Y) \to (X, \Delta + \mathrm{lct}(X, \Delta; \mathfrak{a}) \cdot \mathfrak{a}),$$

where $\Delta_Y = f_*^{-1}\Delta + \mathrm{Ex}(f)$ and for any component $E_i \subset \mathrm{Ex}(f)$ we have

$$A_{X,\Delta}(E) = \mathrm{lct}(X, \Delta; \mathfrak{a}) \cdot \mathrm{mult}_E\, f^*\mathfrak{a}.$$

There is a natural inclusion $\mathfrak{D}(\Delta_Y) \subset \mathrm{Val}^{=1}_{X,x}$, and using a similar argument as in Li and Xu [2014], we can show that there exists a Kollár component S whose rescaling in $\mathrm{Val}^{=1}_{X,x}$ contained in $\mathfrak{D}(\Delta_Y)$ satisfies that

$$\widehat{\mathrm{vol}}(\mathrm{ord}_S) = \mathrm{vol}^{\mathrm{loc}}(-A_{X,\Delta}(S) \cdot S) \le \mathrm{vol}^{\mathrm{loc}}(-K_Y - \Delta_Y) \le \mathrm{mult}(\mathfrak{a}) \cdot \mathrm{lct}^n(X, \Delta; \mathfrak{a}).$$

Then (1) implies that

$$\mathrm{vol}(x, X) = \inf\{\widehat{\mathrm{vol}}(\mathrm{ord}_S) |\ S \text{ is a Kollár component}\}.$$

We can also show that if a minimiser is a divisor then it is indeed a Kollár component (this is proved independently in Blum [2016]).

Moreover, if $x \in (X, \Delta)$ admits a torus group T-action, then by degenerating to the initial ideals, as the colengths are preserved and the log canonical thresholds may only decrease, the right hand side of (1) can be replaced by all T-equivariant ideals. Moreover, equivariant MMP allows us to make all the above data Y and S T-equivariant.

Step 2: In this step, we show that if a minimiser v is quasi-monomial such that $R_0 = \mathrm{gr}_v(R)$ is finitely generated, then the degeneration pair $(X_0 =_{\mathrm{defn}} \mathrm{Spec}(R_0), D_0)$ is klt. After Step 1, this is easy in the case of Theorem 4.6, as the Kollár component is klt. To treat the higher rank case in Theorem 4.7, we verify two ingredients: first we show that a rescaling of ord_{S_i} for the approximating sequence of S_i in Step 1 can be all chosen in the dual complex of a fixed model considered as a subspace of $\mathrm{Val}^{=1}_{X,x}$; then we show as $\mathrm{gr}_v(R)$ is finitely generated, for any i sufficiently large, $\mathrm{gr}_v(R) \cong \mathrm{gr}_{\mathrm{ord}_{S_i}}(R)$. This immediately implies that (X_0, D_0) is the same as the corresponding cone $C(S_0, \Delta_{S_0})$ over the Kollár component S_0, and then we conclude it is klt as before.

Step 3: To proceed we need to establish properties of a general log Fano cone (X_0, Δ_0, ξ_v) and show that the corresponding valuation v is a minimiser if and only if (X_0, Δ_0, ξ_v) is K-semistable. First assume (X_0, Δ_0, ξ_v) is K-semistable, then by Step 1, it suffices to show that for any T-equivariant Kollár component S, $\widehat{\mathrm{vol}}(\mathrm{ord}_S) \geq \widehat{\mathrm{vol}}(v)$. In fact, for any such S, it induces a special degeneration of (X_0, Δ_0, ξ_v) to (Y, Δ_Y, ξ_Y) admitting a $((\mathbb{C}^*)^r \times \mathbb{C}^*)$-action and a new rational vector $\eta_S \in N \oplus \mathbb{Z}$ corresponding to the $\mathbb{C}^*$-action on the special fiber induced by the degeneration. Then an observation going back to Martelli, Sparks, and Yau [2008] says that

$$\frac{d\, \widehat{\mathrm{vol}}(\xi_Y + t \cdot \xi_S)}{dt} = \mathrm{Fut}(Y, \Delta_Y, \xi_Y; \xi_S) \geq 0.$$

Here the generalised Futaki invariant $\mathrm{Fut}(Y, \Delta_Y, \xi_Y; \xi_S)$ is defined in Collins and Székelyhidi [2015, p. 2.2], and then the last inequality comes from the K-semistability assumption. It is also first observed in Martelli, Sparks, and Yau [2008] that the normalised volume function $\widehat{\mathrm{vol}}$ is convex on the space of valuations $\{\xi_v \mid v \in \mathfrak{t}^+_{\mathbb{R}}\}$. Thus by restricting the function on the ray $\xi_Y + t \cdot \xi_S$ $(t \geq 0)$ and applying the convexity, we conclude that

$$\widehat{\mathrm{vol}}_{X_0}(\mathrm{ord}_S) = \lim_{t\to\infty} \widehat{\mathrm{vol}}_Y(\xi_Y + t \cdot \xi_S) \geq \widehat{\mathrm{vol}}_Y(\xi_Y) = \widehat{\mathrm{vol}}_{X_0}(\xi_v).$$

Reversing the argument, one can show that if v is a minimiser of $\widehat{\mathrm{vol}}$ for a log Fano cone singularity (X_0, Δ_0, ξ_v), then for any special degeneration with the same notation as above, we have $\mathrm{Fut}(Y, \Delta_Y, \xi_Y; \xi_S) \geq 0$.

Step 4: An consequence of Step 3 is that for a valuation v on X such that the degeneration (X_0, Δ_0, ξ_v) is K-semistable, since the degeneration to the initial ideal argument implies that $\mathrm{vol}(x, X) \geq \mathrm{vol}(o, X_0)$, then

$$\widehat{\mathrm{vol}}_X(v) = \widehat{\mathrm{vol}}_{X_0}(\xi_v) = \mathrm{vol}(o, X_0)$$

is equal to $\mathrm{vol}(x, X)$.

Step 5: Then we proceed to show that if a log Fano cone (X_0, Δ_0, ξ_v) comes from a degeneration of a minimiser as in Step 2, then it is K-semistable. If not, by Step 3, we can find a degeneration (Y, Δ_Y, ξ_Y) induced by an equivariant Kollár component S with $\widehat{\mathrm{vol}}_Y(\mathrm{ord}_S) < \widehat{\mathrm{vol}}_Y(\xi_Y) = \widehat{\mathrm{vol}}_{X_0}(\xi_v)$. Then arguments similar to Anderson [2013, Section 5] show we can construct a degeneration of (X, Δ) to (Y, Δ_Y) and a family of valuations $v_t \in \mathrm{Val}_{X,x}$ for $t \in [0, \epsilon]$ (for some $0 < \epsilon \ll 1$), with the property that

$$\widehat{\mathrm{vol}}_X(v_t) = \widehat{\mathrm{vol}}_Y(\xi_Y + t \cdot \xi_S) < \widehat{\mathrm{vol}}_Y(\xi_Y) = \widehat{\mathrm{vol}}_{X_0}(\xi_v) = \widehat{\mathrm{vol}}_X(v),$$

where for the second inequality, we use again the fact that $\widehat{\mathrm{vol}}_Y(\xi_Y + t \cdot \xi_S)$ is a convex function. But this is a contradiction.

Step 6: Now we turn to the uniqueness. In this step, we show this for a K-semistable Fano cone singularity (X_0, Δ_0, ξ_v). In fact, for any T-equivariant valuation μ, we can connect ξ_v and μ by a path μ_t such that $\mu_0 = \xi_v$ and $\mu_1 = \mu$. A Newton-Okounkov body type construction (similar to Kaveh and Khovanskii [2014]) can interpret the volumes $\widehat{\mathrm{vol}}(\mu_t)$ to be the volumes of the regions $\mathfrak{A}_t$ contained in the convex cone $\mathfrak{C}$ cut out by a hyperplane H_t passing through a given vector inside $\mathfrak{C}$. Then we conclude by the fact in the convex geometry which says that such a function $f(t) = \mathrm{vol}(\mathfrak{A}_t)$ is strictly convex. Thus it has a unique minimiser, which is ξ_v by Step 3.

Step 7: The last step is to prove the uniqueness in general, under the assumption that it admits a degeneration (X_0, Δ_0, ξ_v) given by a K-semistable minimiser v. For another quasi-monomial minimiser v' of rank r', by a combination of the Diophantine approximation and an MMP construction including the application of ACC of log canonical thresholds (see Section 3), we can obtain a model $f : Z \to X$ which extracts r' divisors E_i $(i = 1, \dots, r')$ such that $(Z, \Delta_Z =_{\mathrm{defn}} \sum E_i + f_*^{-1}\Delta)$ is log canonical. Moreover, the quasi-monomial valuation v' can be computed at the generic point of a component of the intersection of E_i, along which (Z, Δ_Z) is toroidal. Then with the help of the MMP, a careful analysis can show $Z \to X$ degenerates to a birational morphism $Z_0 \to X_0$. Moreover, there exists a quasi-monomial valuation w computed on Y_0 which can be considered as a degeneration of v' with

$$\widehat{\mathrm{vol}}_{X_0}(w) = \widehat{\mathrm{vol}}_X(v') = \widehat{\mathrm{vol}}_X(v) = \widehat{\mathrm{vol}}_{X_0}(\xi_v).$$

Thus $w = \xi_v$ by Step 5 after a rescaling. Since $w(\mathbf{in}(f)) \geq v'(f)$ and $\mathrm{vol}(w) = \mathrm{vol}(v')$, we may argue this implies $\xi_v(\mathbf{in}(f)) = v'(f)$. Therefore, v' is uniquely determined by ξ_v. □

Weaker than Theorem 4.6, in Theorem 4.7 we can not show the finite generation of $\mathrm{gr}_v(R)$, thus we have to post it as an assumption. This is due to the fact that unlike in the divisorial case where the construction of Kollár component provides a satisfying birational model to understand ord_S, for a quasi-monomial valuation of higher rank, the auxiliary models (see Step 1 and 6 in the above proof) we construct are less canonical. Moreover, compared to the statement in Conjecture 4.4, it remains wide open to verify that the minimiser is always quasi-monomial.

One of the main applications of Theorem 4.6 and Theorem 4.7 is to address Donaldson–Sun's conjecture in Donaldson and Sun [2017] on the algebraicity of the construction of the metric tangent cone, which can be considered as a local analogue of Donaldson and Sun [2014], G. Tian [2013], and Li, Wang, and Xu [2014]. More precisely, it was proved that the Gromov-Hausdorff limit of a sequence of Kähler-Einstein metric Fano varieties is a Fano variety X_∞ with klt singularities (see Donaldson and Sun [2014] and G. Tian [2013]). And to understand the metric structure near a singularity $x \in X_\infty$, we need to understand its metric tangent cone C (cf. Cheeger, Colding, and G. Tian [2002]). In the work Donaldson and Sun [2017], a description of C was given by a two-step degeneration process: first there is a valuation v on $\mathrm{Val}_{X_\infty,x}$ whose associated graded ring induces a degeneration of $x \in X_\infty$ to $o \in M$; then there is a degeneration of Fano cone from $o \in M$ to $o' \in C$. In Donaldson-Sun's definitions of M and C, they used the local metric structure around $x \in X_\infty$. However, they conjectured that both M and C only depend on the underlying algebraic structure of the germ $x \in X_\infty$. Built on the previous works of Li [2015a], Li and Liu [2016], and Li and Xu [2016], we answer the first part of their conjecture affirmatively, which says M is determined by the algebraic structure of the germ $x \in X_\infty$. We achieve this by showing that v is a K-semistable valuation in $\mathrm{Val}_{X,x}$ and such a K-semistable valuation is unique up to rescaling.

Theorem 4.8 (Li and Xu [2017]). *The valuation v is the unique minimiser (up to scaling) of $\widehat{\mathrm{vol}}$ in all quasi-monomial valuations in $\mathrm{Val}_{X_\infty,x}$.*

Proof. From the results proved in Donaldson and Sun [2017], we can verify that $o \in (W, \xi_v)$ is a K-semistable Fano cone singularity, which exactly means v is a K-semistable valuation. Thus v is a minimiser of $\widehat{\mathrm{vol}}$ by the last statement of Theorem 4.7. Then up to rescaling, v is the unique quasi-monomial minimiser again by Theorem 4.7. □

We expect that the tools we developed, especially those on equivariant K-stability, are enough to solve the second part of Donaldson-Sun's conjecture, i.e. to confirm the metric tangent cone C only depends on the algebraic structure of $x \in X_\infty$.

4.2 The volume of a klt singularity.

As $\mathrm{vol}(x, X)$ carries deep information on the singularity $x \in X$, calculating this number consists an important part of the theory. It also has applications to global questions. We discuss some related results and questions in this section.

In general, it could be difficult to compute $\mathrm{vol}(x, X)$. Even for the smooth point $x \in \mathbb{C}^n$, knowing $\mathrm{vol}(x, \mathbb{C}^n)$ (which is, not surprisingly, equal to n^n) involves highly nontrivial arguments. An illuminating example is the following.

Example 4.9 (Li [2015a], Li and Liu [2016], and Li and Xu [2016])**.** A $\mathbb{Q}$-Fano variety is K-semistable if and only if for the cone $C = C(X, -rK_X)$, the canonical valuation v obtained by blowing up the vertex o is a minimiser.

On one hand, this means that finding out the minimiser is in general at least as hard as testing the K-semistablity of (one dimensional lower) Fano varieties, which has been known to be a challenging question; on the other hand, this sheds new light on the question of testing K-stability. For example, using properties of degenerating ideals to their initials, we can prove that for a klt Fano variety X with a torus group T-action, to test the K-semistability of X it suffices to test on T-equivariant special test configurations (see Li and Xu [2016]).

The Stable Degeneration Conjecture 4.4 implies many properties of $\mathrm{vol}(x, X)$. The first one we want to discuss is a finite degree multiplication formula.

Conjecture 4.10. *If $\pi : x_1 \in (X_1, \Delta_1) \to x_2 \in (X_2, \Delta_2)$ is a finite dominant morphism between klt singularities such that $\pi^*(K_{X_2} + \Delta_2) = K_{X_1} + \Delta_1$, then*

$$\deg(\pi) \cdot \mathrm{vol}(x_2, X_2) = \mathrm{vol}(x_1, X_1).$$

This can be easily reduced to the case that the finite covering $X_1 \to X_2$ is Galois, and we denote the Galois group by G. Then it suffices to show that the minimiser of X_1 is G-equivariant, which is implied by the uniqueness claim in Conjecture 4.4. Conjecture 4.10 is verified in Li and Xu [2017] for $x \in X_\infty$ where X_∞ is a Gromov-Hausdorff limit of Kähler-Einstein Fano manifolds. Since any point will have its volume less or equal to n^n (see Liu and Xu [2017, Appendix]), Conjecture 4.10 implies that for a klt singularity $x \in (X, \Delta)$,

$$\mathrm{vol}(x, X) \leq n^n / |\hat{\pi}_1^{\mathrm{loc}}(x, X)|, \tag{2}$$

where the finiteness of $\hat{\pi}_1^{\mathrm{loc}}(x, X)$ is proved in Xu [2014].

Combining Conjecture 4.4 with the well known speculation that K-semistable is a Zariski open condition, we also have the following conjecture.

Conjecture 4.11. *Given a klt pair (X, Δ), then the function $\mathrm{vol}(x, X)$ is a constructible function, i.e. we can stratify X into constructible sets $X = \sqcup_i S_i$, such that for any i, $\mathrm{vol}(x, X)$ takes a constant value for all $x \in S_i$.*

A degeneration argument in Liu [2017] implies that the volume function should be lower semi-continuous. A special case we know is that the volume of any n-dimensional klt non-smooth point is always less than n^n (see Liu and Xu [2017, Appendix]).

Finally, we discuss some applications of the volume of singularities to K-stability of Fano varieties. A useful formula connecting local and global geometries is the following.

Theorem 4.12 (Fujita [2015] and Liu [2016])**.** *If X is a K-semistable $\mathbb{Q}$-Fano variety, then for any point $x \in X$, we have*

$$\operatorname{vol}(x, X) \geq \left(\frac{n+1}{n}\right)^n (-K_X)^n. \tag{3}$$

So if we can bound the type of klt singularities from the lower bound of their volumes, then we can restrict the type of singularities appearing on a K-semistable $\mathbb{Q}$-Fano variety with a given volume. In particular, this applies to the Gromov-Hausdorff limit X_∞ of a sequence of Kähler-Einstein Fano manifolds X_i (with a large volume of $-K_{X_i}$). If the restriction is sufficiently effective, then X_∞ would appear in an explicit simple ambient space on which we can carry out the orbital geometry calculation to identify X_∞ by showing all other possible limits are K-unstable.

For instance, by revisiting the classification results of three dimensional singularities, we show that $\operatorname{vol}(x, X) \leq 16$ if $x \in X$ is singular and the equality holds if and only if $x \in X$ is a rational double point (see Liu and Xu [2017]). As a consequence, we could solve the question on the existence of Kähler-Einstein metrics for cubic threefolds.

Corollary 4.13 (Liu and Xu [ibid.])**.** *GIT polystable (resp. semistable) cubic threefolds are K-polystable (resp. K-semistable). In particular, all GIT polystable cubic threefolds, including every smooth one, admit Kähler-Einstein metrics.*

References

D. Abramovich, K. Karu, K. Matsuki, and J. Włodarczyk (2002). "Torification and factorization of birational maps". *J. Amer. Math. Soc.* 15.3, pp. 531–572 (cit. on p. 829).

V. Alexeev (1994). "Boundedness and K^2 for log surfaces". *Internat. J. Math.* 5.6, pp. 779–810 (cit. on pp. 834, 835).

D. Anderson (2013). "Okounkov bodies and toric degenerations". *Math. Ann.* 356.3, pp. 1183–1202 (cit. on p. 843).

U. Angehrn and Y. Siu (1995). "Effective freeness and point separation for adjoint bundles". *Invent. Math.* 122.2, pp. 291–308 (cit. on p. 835).

V. Arnol'd, S. Guseĭn-Zade, and A. Varchenko (1985). *Singularities of differentiable maps. Vol. I*. Vol. 82. Monographs in Mathematics. Birkhäuser Boston, Inc., Boston, MA, pp. xi+382. MR: 777682 (cit. on p. 832).

C. Birkar (2016). "Singularities of linear systems and boundedness of Fano varieties". arXiv: 1609.05543 (cit. on p. 835).

C. Birkar, P. Cascini, C. Hacon, and J. McKernan (2010). "Existence of minimal models for varieties of log general type". *J. Amer. Math. Soc.* 23.2, pp. 405–468 (cit. on pp. 829, 833, 836).

H. Blum (2016). "Existence of Valuations with Smallest Normalized Volume". To appear in Compos. Math. arXiv: 1606.08894 (cit. on pp. 838, 841).

J. Cheeger, T. Colding, and G. Tian (2002). "On the singularities of spaces with bounded Ricci curvature". *Geom. Funct. Anal.* 12.5, pp. 873–914 (cit. on p. 844).

X. Chen, S. Donaldson, and S. Sun (2015). "Kähler-Einstein metrics on Fano manifolds. I: Approximation of metrics with cone singularities, II: Limits with cone angle less than 2π, III: Limits as cone angle approaches 2π and completion of the main proof." *J. Amer. Math. Soc.* 28.1, pp. 183–197, 199–234, 235–278 (cit. on p. 838).

T. Collins and G. Székelyhidi (2015). "Sasaki-Einstein metrics and K-stability". arXiv: 1512.07213 (cit. on pp. 840, 842).

L. Denef and J. Loeser (2001). "Geometry on arc spaces of algebraic varieties". In: *European Congress of Mathematics, Vol. I (Barcelona, 2000)*. Vol. 201. Progr. Math. Birkhäuser, Basel, pp. 327–348 (cit. on p. 830).

S. Donaldson (2001). "Scalar curvature and projective embeddings. I". *J. Differential Geom.* 59.3, pp. 479–522 (cit. on p. 840).

S. Donaldson and S. Sun (2014). "Gromov-Hausdorff limits of Kähler manifolds and algebraic geometry". *Acta Math.* 213.1, pp. 63–106 (cit. on p. 844).

– (2017). "Gromov-Hausdorff limits of Kähler manifolds and algebraic geometry, II". *J. Differential Geom.* 107.2, pp. 327–371 (cit. on p. 844).

L. Ein, R. Lazarsfeld, and K. Smith (2003). "Uniform approximation of Abhyankar valuation ideals in smooth function fields". *Amer. J. Math.* 125.2, pp. 409–440 (cit. on pp. 837, 839).

T. de Fernex, L. Ein, and M. Mustaţă (2010). "Shokurov's ACC conjecture for log canonical thresholds on smooth varieties". *Duke Math. J.* 152.1, pp. 93–114 (cit. on pp. 832, 833).

T. de Fernex, J. Kollár, and C. Xu (2017). "The dual complex of singularities". In: *Higher dimensional algebraic geometry–in honour of Professor Yujiro Kawamata's sixtieth birthday*. Advanced Studies in Pure Mathematics 74. Math. Soc. Japan, Tokyo, pp. 103–130. arXiv: 1212.1675 (cit. on pp. 828, 829, 836).

O. Fujino (2011). "Semi-stable minimal model program for varieties with trivial canonical divisor". *Proc. Japan Acad. Ser. A Math. Sci.* 87.3, pp. 25–30 (cit. on p. 830).

K. Fujita (2015). "Optimal bounds for the volumes of Kähler-Einstein Fano manifolds". To appear in Amer. J. Math. arXiv: 1508.04578 (cit. on pp. 838, 846).

M. Gross and B. Siebert (2011). "From real affine geometry to complex geometry". *Ann. of Math. (2)* 174.3, pp. 1301–1428 (cit. on p. 830).

C. Hacon and J. McKernan (2006). "Boundedness of pluricanonical maps of varieties of general type". *Invent. Math.* 166.1, pp. 1–25 (cit. on p. 835).

C. Hacon, J. McKernan, and C. Xu (2013). "On the birational automorphisms of varieties of general type". *Ann. of Math. (2)* 177.3, pp. 1077–1111 (cit. on p. 835).

– (2014). "ACC for log canonical thresholds". *Ann. of Math.* 180 (2), pp. 523–571 (cit. on pp. 832–835).

– (2016). "Boundedness of varieties of log general type". In: *To appear in Proceedings of 2015 Summer Research Institute on Algebraic Geometry* (cit. on p. 834).

C. Hacon and C. Xu (2013). "Existence of log canonical closures". *Invent. Math.* 192.1, pp. 161–195 (cit. on p. 829).

M. Jonsson and M. Mustaţă (2012). "Valuations and asymptotic invariants for sequences of ideals". *Ann. Inst. Fourier (Grenoble)* 62.6, pp. 2145–2209 (cit. on pp. 837, 839).

K. Kaveh and A. Khovanskii (2014). "Convex bodies and multiplicities of ideals". *Proc. Steklov Inst. Math.* 286.1, pp. 268–284 (cit. on p. 843).

J. Kollár (1994). "Log surfaces of general type; some conjectures", pp. 261–275 (cit. on p. 834).

– (1997). "Singularities of pairs". In: *Algebraic geometry—Santa Cruz 1995*. Vol. 62. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 221–287 (cit. on pp. 832, 835).

– (2013a). "Moduli of varieties of general type". In: *Handbook of moduli. Vol. II*. Vol. 25. Adv. Lect. Math. (ALM). Int. Press, Somerville, MA, pp. 131–157. MR: 3184176 (cit. on p. 834).

– (2013b). *Singularities of the minimal model program*. Vol. 200. Cambridge Tracts in Mathematics. With a collaboration of Sándor Kovács. Cambridge: Cambridge University Press (cit. on pp. 827, 828, 833).

J. Kollár and S. Kovács (2010). "Log canonical singularities are Du Bois". *J. Amer. Math. Soc.* 23.3, pp. 791–813 (cit. on p. 831).

J. Kollár and S. Mori (1998). *Birational geometry of algebraic varieties*. Vol. 134. Cambridge Tracts in Mathematics. With the collaboration of C. H. Clemens and A. Corti; Translated from the 1998 Japanese original. Cambridge University Press, pp. viii+254 (cit. on pp. 825, 827).

J. Kollár and C. Xu (2016). "The dual complex of Calabi–Yau pairs". *Invent. Math.* 205.3, pp. 527–557 (cit. on p. 831).

M. Kontsevich and Y. Soibelman (2001). "Homological mirror symmetry and torus fibrations". In: *Symplectic geometry and mirror symmetry (Seoul, 2000)*. World Sci. Publ., River Edge, NJ, pp. 203–263 (cit. on p. 830).

Robert Lazarsfeld (2004). *Positivity in algebraic geometry. II*. Vol. 49. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics. Springer-Verlag, Berlin, pp. xviii+385 (cit. on p. 838).

C. Li (2015a). "K-semistability is equivariant volume minimization". To appear in Duke Math. J. arXiv: 1512.07205 (cit. on pp. 840, 844, 845).

– (2015b). "Minimizing normalized volumes of valuations". To appear in Math. Zeit. arXiv: 1511.08164 (cit. on pp. 837, 838).

C. Li and Y. Liu (2016). " Kähler-Einstein metrics and volume minimization". arXiv: 1602.05094 (cit. on pp. 844, 845).

C. Li, X. Wang, and C. Xu (2014). " On proper moduli space of smoothable Kähler-Einstein Fano varieties". arXiv: 1411.0761v3 (cit. on pp. 838, 844).

C. Li and C. Xu (2014). "Special test configuration and K-stability of Fano varieties". *Ann. of Math. (2)* 180.1, pp. 197–232 (cit. on pp. 838, 841).

– (2016). " Stability of Valuations and Kollár Components". arXiv: 1604.05398 (cit. on pp. 840, 844, 845).

– (2017). " Stability of Valuations: Higher Rank". arXiv: 1707.05561 (cit. on pp. 838–841, 844, 845).

Y. Liu (2016). "The volume of singular Kähler-Einstein Fano varieties". To appear in Compos. Math. arXiv: 1605.01034 (cit. on pp. 837, 846).

– (2017). "On the semi-continuity problem of normalized volumes of singularities". arXiv: 1711.06962 (cit. on p. 846).

Y. Liu and C. Xu (2017). "K-stability of cubic threefolds". arXiv: 1706.01933 (cit. on pp. 845, 846).

D. Martelli, J. Sparks, and S. Yau (2008). "Sasaki-Einstein manifolds and volume minimisation". *Comm. Math. Phys.* 280.3, pp. 611–673 (cit. on pp. 839, 842).

J. McKernan and Yuri Prokhorov (2004). "Threefold thresholds". *Manuscripta Math.* 114.3, pp. 281–304 (cit. on p. 835).

J. Milnor (1968). *Singular points of complex hypersurfaces*. Annals of Mathematics Studies, No. 61. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, pp. iii+122 (cit. on p. 829).

D. Mumford (1961). "The topology of normal singularities of an algebraic surface and a criterion for simplicity". *Inst. Hautes Études Sci. Publ. Math.* 9, pp. 5–22 (cit. on pp. 828, 829).

M. Mustaţă and J. Nicaise (2015). "Weight functions on non-Archimedean analytic spaces and the Kontsevich-Soibelman skeleton". *Algebr. Geom.* 2.3, pp. 365–404 (cit. on p. 830).

J. Nicaise and C. Xu (2016a). "Poles of maximal order of motivic zeta function". *Duke Math. J.* 165.2, pp. 217–243 (cit. on p. 830).

J. Nicaise and C. Xu (2016b). "The essential skeleton of a degeneration of algebraic varieties". *Amer. J. Math.* 138.6, pp. 1645–1667 (cit. on pp. 830, 831).

Y. Odaka (2013). "The GIT stability of polarized varieties via discrepancy". *Ann. of Math. (2)* 177.2, pp. 645–661 (cit. on p. 838).

S. Payne (2013). "Boundary complexes and weight filtrations". *Michigan Math. J.* 62.2, pp. 293–322 (cit. on p. 829).

V. V. Shokurov (1992). "Three-dimensional log perestroikas". *Izv. Ross. Akad. Nauk Ser. Mat.* 56.1, pp. 105–203 (cit. on p. 832).

A. Strominger, S. Yau, and E. Zaslow (1996). "Mirror symmetry is T-duality". *Nuclear Phys. B* 479.1-2, pp. 243–259 (cit. on p. 830).

S. Takayama (2006). "Pluricanonical systems on algebraic varieties of general type". *Invent. Math.* 165.3, pp. 551–587 (cit. on p. 835).

B. Teissier (2003). "Valuations, deformations, and toric geometry". In: *Valuation theory and its applications, Vol. II (Saskatoon, SK, 1999)*. Vol. 33. Fields Inst. Commun. Amer. Math. Soc., Providence, RI, pp. 361–459 (cit. on p. 839).

Gang Tian (1997). "Kähler-Einstein metrics with positive scalar curvature". *Invent. Math.* 130.1, pp. 1–37 (cit. on p. 840).

– (2013). "Partial C^0-estimate for Kähler-Einstein metrics". *Commun. Math. Stat.* 1.2, pp. 105–113 (cit. on p. 844).

– (2015). "K-stability and Kähler-Einstein metrics". *Comm. Pure Appl. Math.* 68.7, pp. 1085–1156 (cit. on p. 838).

Z. Tian and C. Xu (2016). " Finiteness of fundamental groups". *Compos. Math.* 153.2, pp. 257–273 (cit. on p. 837).

H. Tsuji (2007). "Pluricanonical systems of projective varieties of general type. II". *Osaka J. Math.* 44.3, pp. 723–764 (cit. on p. 835).

C. Xu (2014). "Finiteness of algebraic fundamental groups". *Compos. Math.* 150.3, pp. 409–414 (cit. on pp. 831, 836, 837, 845).

Received 2017-12-05.

CHENYANG XU (许晨阳)
BEIJING INTERNATIONAL CENTER FOR MATHEMATICAL RESEARCH
BEIJING 100871
CHINA
cyxu@math.pku.edu.cn

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (849–876)

HODGE THEORY AND CYCLE THEORY OF LOCALLY SYMMETRIC SPACES

NICOLAS BERGERON

Abstract

We discuss several results pertaining to the Hodge and cycle theories of locally symmetric spaces. The unity behind these results is motivated by a vague but fruitful analogy between locally symmetric spaces and projective varieties.

1 Introduction

Locally symmetric spaces are complete Riemannian manifolds locally modeled on certain homogeneous spaces. Their holonomy groups are typically smaller than SO_n – the holonomy group of a generic Riemannian manifold – and there are invariant tensors on the tangent space that are preserved by parallel transport. It was first observed by Chern [1957] that Hodge theory can be used to promote these local algebraic structures to structures that exist on the cohomology groups of locally symmetric spaces. This is very similar to what happens for compact Kähler manifolds. In fact the analogy between locally symmetric spaces and Kähler manifolds – or rather complex projective varieties – is a fruitful one in many aspects. In this report we shall discuss several instances of this analogy. We don't give proofs, we only state recent results that illustrate various items of the following dictionary.

	Projective varieties $V \subset \mathbb{P}^n$	**Locally symmetric spaces $\Gamma \backslash S$**
Complexity	degree	volume
Hodge theory	Hodge-Lefschetz decomposition	Matsushima's fomula and $(\mathfrak{g}, K)$-cohomology
Cycles	Algebraic cycles (Intersections of) Hyperplane sections	Modular cycles Sums of Hecke translates of modular cycles
Cohomology	Kodaira vanishing Theorem Lefschetz hyperplane Theorem Hodge Conjecture	Vanishing theorems using spinors Automorphic Lefschetz' properties Hodge type theorems

MSC2010: primary 11F75; secondary 14C30, 14G35, 57T25, 11F70, 22E40.

The following result (see Theorem 8 below), jointly obtained with Millson and Moeglin, shows that the right side of the above dictionary may eventually shed some light on the left (more classical) side.

Theorem. *On a projective unitary Shimura variety uniformized by the complex n-ball, any Hodge* (r,r)*-class with* $r \in [0,n] \backslash]\frac{n}{3}, \frac{2n}{3}[$ *is algebraic.*

Context. There has been a great deal of work on the cohomology of locally symmetric spaces. This involves methods from geometry, analysis and number theory. We note in particular that related topics have been discussed Harris [2014], Venkataramana [2010], and Speh [2006] in the last three ICMs. Indeed, Harris [2014] contains an overview of the program for analyzing cohomology of Shimura varieties developed by Langlands and Kottwitz. It aims at attaching Galois representations to the corresponding cohomology classes. Our point of view is closer to Venkataramana [2010] and Speh [2006] that discuss conjectures that naturally fit into the above dictionnary. The latter has been very much influenced by former works of Oda, Venkataramana, Harris-Li discussed in Venkataramana [2010]. We also have borrowed some expository ideas from §3 of Venkatesh Takagi lectures Venkatesh [2017].

2 Locally symmetric spaces

2.1 Symmetric spaces. A *symmetric space* is a Riemannian manifold whose group of symmetries contains an inversion symmetry about every point. We will be mainly concerned with symmetric spaces of non-compact type. Such a space S is associated to a connected center-free semi-simple Lie group G without compact factor. As a manifold S is the quotient G/K of G by a maximal compact subgroup $K \subset G$; it is known that all such K are conjugate inside G. One may easily verify that G preserves a Riemannian metric on S. Unless otherwise specified our symmetric spaces S will always be assumed to be of non-compact types.

For example, if $G = \mathrm{PSL}_2(\mathbb{R})$, we can take $K = \mathrm{PSO}_2$, and the associated symmetric space $S = G/K$ can be identified with the Poincaré upper-half plane $\mathbb{H}^2 = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ and the action of G is by fractional linear transformations; it preserves the standard hyperbolic metric $|dz|^2/\mathrm{Im}(z)^2$.

If $G = \mathrm{PSL}_2(\mathbb{C})$, we can take $K = \mathrm{PSU}_2$, and the associated symmetric space S can be identified with the three-dimensional hyperbolic space $\mathbb{H}^3$.

We shall be particularly concerned with cases where G is either a unitary group $\mathrm{PU}(p,q)$ or an orthogonal group $\mathrm{SO}_0(p,q)$. Thanks to special isomorphisms between low dimensional Lie groups the symmetric spaces associated to $\mathrm{PU}(1,1)$ and $\mathrm{SO}_0(2,1)$ are both isometric to the Poincaré upper-half plane $\mathbb{H}^2$ and the symmetric space associated to $SO_0(3,1)$ is isometric to the three-dimensional hyperbolic space $\mathbb{H}^3$.

Another important case to consider is that of $G = PSL_n(\mathbb{R})$. Then we can take $K = PSO_n$, and the symmetric space S can be identified with the space of positive definite, symmetric, real valued $n \times n$ matrices A with $\det(A) = 1$, with metric given by $\mathrm{trace}(A^{-1}dA)^2$.

2.2 Locally symmetric spaces. Locally symmetric spaces are complete Riemannian manifolds locally modeled on some symmetric space S with changes of charts given by restrictions of elements of G. It is known that all such manifolds are isometric to quotients $\Gamma\backslash S$ of S by some discrete torsion-free subgroup $\Gamma \subset G$. One may measure the complexity of a locally symmetric space by considering its volume, or equivalently the volume of a fundamental domain for the action of Γ on S. We shall be only concerned with locally symmetric spaces of finite volume; the group Γ is then a lattice in G – in many cases we shall even restrict to compact locally symmetric spaces.

By a general theorem of Borel, any symmetric space S admits a compact manifold quotient (S-manifold) $\Gamma\backslash S$. In Borel's construction Γ is a *congruence arithmetic group*. For our purpose let us define these groups as those obtained by taking a semi-simple algebraic $\mathbb{Q}$-group $\mathbf{H} \subset SL_N$, and taking

$$\{h \in \mathbf{H}(\mathbb{Q}) \ : \ h \text{ has integral entries}\}. \tag{2-1}$$

Each such group is contained in an ambient Lie group, namely the real points of $\mathbf{H}$. If $\mathbf{H}(\mathbb{R})$ is isogeneous to $G \times$ (compact) the projection on the first factor maps the discrete subgroup (2-1) onto a lattice Γ in G. If the compact factor in $\mathbf{H}(\mathbb{R})$ is non-trivial then Γ is necessarily co-compact in G. Finally, replacing the discrete group (2-1) by its intersection with the kernel of a reduction mod ℓ map $SL_N(\mathbb{Z}) \to SL_N(\mathbb{Z}/\ell\mathbb{Z})$, one can obtain a torsion-free lattice Γ. We refer to the corresponding locally symmetric spaces $\Gamma\backslash S$ as *congruence arithmetic*.

2.3 Examples. Locally symmetric spaces play a central role in geometry. Here are some important examples:

Shimura varieties. These appear in algebraic geometry as moduli spaces of certain types of Hodge structures. E.g. for all g the moduli space $\mathcal{K}_g$ of genus g quasi-polarized K3 surfaces identifies with a locally symmetric space associated to $G = SO_0(2,19)$. Shimura varieties themselves are quasi-projective varieties. They play an important role in number theory through Langlands' program.

Complex ball quotients. By Yau's solution to the Calabi conjecture, complex algebraic surfaces whose Chern numbers satisfy $c_1^2 = 3c_2$ are quotients of the unit ball in $\mathbb{C}^2$ by a torsion-free co-compact lattice in $\mathrm{PU}(2,1)$. Most famously, this includes the classification of fake projective planes by Klingler [2003], Prasad and Yeung [2009] and Cartwright and Steger [2010]. Picard [1881], Deligne and Mostow [1986] and Thurston [1998] give many examples of ball quotients coming from natural moduli problems. Congruence arithmetic ball quotients are particular Shimura varieties.

Hyperbolic manifolds. In dimension 3, according to Thurston's geometrization conjecture, proved by Perelman, a 'generic' manifold is hyperbolic. More generally, Gromov theory of δ-hyperbolic groups suggest that negative curvature is 'quite generic.' However, at least in dimension ≥ 5, all known (to the author) constructions of closed manifolds that can carry a negatively curved metric are essentially obtained by rather simple surgeries on locally symmetric manifolds. These spaces therefore form a fundamental family of examples in geometry and more generally play a crucial role in geometric group theory.

Teichmüller spaces of flat unimodular metrics on tori $\mathbb{R}^n/\mathbb{Z}^n$. These are locally symmetric spaces associated to $\mathrm{PSL}_n(\mathbb{R})$. Their cohomology groups are very tightly bound to algebraic K-theory. In particular this viewpoint quite naturally leads to the famous regulator of Borel.

2.4 Notation. We have already defined $K \subset G$ and the associated Riemannian symmetric space $S = G/K$. Let $\mathfrak{g}$ be the complexified Lie algebra of G and let G^c be a compact form of G. Let $S^c = G^c/K$ be the compact dual of S. Let θ be the Cartan involution of G fixing K and let $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$ be the associated Cartan decomposition. We normalize the Riemannian metric on S^c such that multiplication by i in $\mathfrak{p}$ becomes an isometry $T_{eK}S \to T_{eK}S^c$.

From now on $\Gamma\backslash S$ will denote a finite volume locally symmetric S-manifold. In general we try to reserve n for its real dimension, or complex dimension if S is Hermitian.

We denote by $b_k(M)$ the Betti numbers of a manifold M.

3 Hodge theory

For simplicity, in this section, we will assume that all the locally symmetric spaces $\Gamma\backslash S$ we consider are compact. This excludes some of the important examples mentioned above. However modified versions of the discussion below still apply and we will abusively ignore this issue in the rest of this document.

Being a compact manifold, the quotient $\Gamma\backslash S$ satisfies Poincaré duality. But, as mentioned in the Introduction, the Riemannian manifold $\Gamma\backslash S$ in general has a much smaller holonomy group than SO_n, and one can show that this forces $\Gamma\backslash S$ to satisfy many more constraints, see (3-5) and (3-6). These constraints can be understood in terms of *cohomological representations*, i.e. unitary representations π of G such that the relative Lie algebra cohomology $H^*(\mathfrak{g}, K; \pi)$ is non-zero (see Section 3.2 below).

Since the general setup of $(\mathfrak{g}, K)$-cohomology is rather forbidding we will discuss in more detail two special examples. But first, let us emphasize the analogy with projective manifolds, or rather here with Kähler manifolds.

3.1 Comparison with Kähler manifolds. Hodge theory gives a way to study the cohomology of a closed Riemannian manifold M. Indeed, each class in $H^*(M, \mathbb{C})$ has a *canonical* 'harmonic' representative: a differential form ω that represents this class and is of minimal L^2 norm. Equivalently the form ω is annihilated by the Hodge-Laplace operator Δ. One gets

$$\underbrace{\text{harmonic } k\text{-forms on } M}_{:=\mathcal{H}^k(M)} \xrightarrow{\simeq} H^k(M, \mathbb{C}). \tag{3-1}$$

Suppose furthermore that M is an n-dimensional complex Kähler manifold. Then, its holonomy group is contained in the unitary group $\mathrm{U}_n \subset \mathrm{SO}_{2n}$ and there is an action of $\mathbb{C}^*$ on each tangent space that is preserved by parallel transport. This yields an action of $\mathbb{C}^*$ on differential forms with complex coefficients. A crucial aspect of the theory of Kähler manifolds is that this action preserves harmonic forms. It then follows from Hodge theory that $\mathbb{C}^*$ acts on the cohomology groups and this gives rise to the Hodge decomposition.

3.2 Matsushima's formula. Let us now come back to the case of a compact locally symmetric manifold $\Gamma\backslash S$.

Because the cotangent bundle $T^*(\Gamma\backslash S)$ is isomorphic to the bundle $\Gamma\backslash G \times_K \mathfrak{p}^* \to \Gamma\backslash G/K$, which is associated to the principal K-bundle $K \to \Gamma\backslash G \to \Gamma\backslash S$ and the adjoint representation of K in $\mathfrak{p}^*$, the space of differential k-forms on $\Gamma\backslash S$ can be identified with $\mathrm{Hom}_K(\wedge^k \mathfrak{p}, C^\infty(\Gamma\backslash G))$. The corresponding complex computes the $(\mathfrak{g}, K)$-cohomology of $C^\infty(\Gamma\backslash G)$ – the subspace of smooth vectors in the right (quasi-)regular representation of G in $L^2(\Gamma\backslash G)$. One similarly defines the $(\mathfrak{g}, K)$-cohomology groups $H^\bullet(\mathfrak{g}, K; \pi)$ of any unitarizable $(\mathfrak{g}, K)$-module (π, V_π). By a theorem of Harish-Chandra, the set of equivalence classes of irreducible unitarizable $(\mathfrak{g}, K)$-modules is naturally identified with the set of equivalence classes of irreducible unitary representations of G. In the following, we will abusively use the same notation to denote an irreducible unitary representation and its associated $(\mathfrak{g}, K)$-module of smooth vectors.

The decomposition of $\wedge^\bullet \mathfrak{p}^*$ into irreducible K-modules induces a decomposition of the exterior algebra $\wedge^\bullet T^*(\Gamma\backslash S) = \Gamma\backslash G \times_K \wedge^\bullet(\mathfrak{p}^*)$. This decomposition commutes with the action of the Hodge-Laplace operator, giving birth to a decomposition of the cohomology $H^\bullet(\Gamma\backslash S, \mathbb{C})$ which refines the Hodge decomposition if S is Hermitian symmetric and gives an analogous decomposition of the cohomology in the case S is not Hermitian. In both cases we will call this decomposition of $H^\bullet(\Gamma\backslash S, \mathbb{C})$ the *generalized Hodge decomposition*; it is better understood in terms of cohomological representations through Matsushima's formula:

$$H^\bullet(\Gamma\backslash S, \mathbb{C}) = \bigoplus_\pi m(\pi, \Gamma) H^\bullet(\mathfrak{g}, K; \pi). \tag{3-2}$$

Here the (finite) sum is over (classes of) irreducible unitary representations of G such that $H^\bullet(\mathfrak{g}, K; \pi) \neq 0$ and $m(\pi, \Gamma)$ is the (finite) multiplicity with which π occurs in the quasi-regular representation $L^2(\Gamma\backslash G)$.

Cohomological representations of G are classified in terms of the θ-stable parabolic subalgebras $\mathfrak{q} \subset \mathfrak{g}$. Let $\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{u}$ be the θ-stable Levi decomposition of $\mathfrak{q}$. We have $\mathfrak{u} = \mathfrak{u} \cap \mathfrak{k} \oplus \mathfrak{u} \cap \mathfrak{p}$. Put $R = \dim(\mathfrak{u} \cap \mathfrak{p})$. The line $\wedge^R(\mathfrak{u} \cap \mathfrak{p})$ generates an irreducible representation $V(\mathfrak{q})$ of K in $\wedge^R \mathfrak{p}$.

The classification of unitary irreducible cohomological representations of G associates to each θ-stable parabolic subalgebras $\mathfrak{q} \subset \mathfrak{g}$ a cohomological representation $A_\mathfrak{q}$ characterized by the property that the only irreducible K-representation common to $\wedge^\bullet \mathfrak{p}$ and $A_\mathfrak{q}$ is the representation $V(\mathfrak{q})$. Moreover, every cohomological representation is an $A_\mathfrak{q}$, see Vogan and Zuckerman [1984].

Each $H^\bullet(\mathfrak{g}, K; A_\mathfrak{q})$ identifies with the cohomology – with degree shifted by R – of the *compact* symmetric space associated to a subgroup $L \subset G^c$ with complexified Lie algebra $\mathfrak{l}$. In particular, the component corresponding to the trivial representation of G in (3-2) is isomorphic to $H^\bullet(S^c, \mathbb{C})$. In the Hermitian case we recover Hirzebruch proportionality principle.

If ω belongs to $H^R(\Gamma\backslash S, \mathbb{C})$ and, under the Matsushima decomposition (3-2), lies in the component corresponding to some $A_\mathfrak{q}$ with $R = \dim(\mathfrak{u} \cap \mathfrak{p})$, by analogy with the notion of primitive class in the Hodge-Lefschetz decomposition, we refer to ω as a *strongly primitive* class of type $A_\mathfrak{q}$.

3.3 Two families of examples.

3.3.1 Compact quotients of the symmetric space associated to $\mathrm{PU}(p, q)$**.** Then the holonomy group is contained in $\mathrm{U}_p \times \mathrm{U}_q$ and S^c is the complex Grassmannian $\mathrm{Gr}_p(\mathbb{C}^{p+q})$

of p-planes in $\mathbb{C}^{p+q}$. We first consider the decomposition of $\wedge^\bullet \mathfrak{p}^*$ into irreducible K-modules. The symmetric space S being of Hermitian type, the exterior algebra $\wedge^\bullet \mathfrak{p}$ decomposes as:

$$\wedge^\bullet \mathfrak{p} = \wedge^\bullet \mathfrak{p}' \otimes \wedge^\bullet \mathfrak{p}'' \tag{3-3}$$

where $\mathfrak{p}'$ and $\mathfrak{p}''$ respectively denote the holomorphic and anti-holomorphic tangent spaces. In the case $q = 1$ – then S is the complex ball of dimension p – it is an exercise to check that the decomposition of (3-3) into irreducible modules recovers the usual Lefschetz decomposition. But, in general, the decomposition is much finer, and it is hard to write down the full decomposition of (3-3) into irreducible modules. Indeed: as a representation of $\mathrm{GL}_p(\mathbb{C}) \times \mathrm{GL}_q(\mathbb{C})$ the space $\mathfrak{p}'$ is isomorphic to $V_+ \otimes V_-^*$ where $V_+ = \mathbb{C}^p$ (resp. $V_- = \mathbb{C}^q$) is the standard representation of $\mathrm{GL}_p(\mathbb{C})$ (resp. $\mathrm{GL}_q(\mathbb{C})$) and the decomposition of $\wedge^\bullet \mathfrak{p}'$ is already quite complicated (see Fulton [1997, Equation (19), p. 121]):

$$\wedge^R (V_+ \otimes V_-^*) \cong \bigoplus_{\lambda \vdash R} S_\lambda(V_+) \otimes S_{\lambda^*}(V_-)^*. \tag{3-4}$$

Here we sum over all partitions of R (equivalently Young diagrams of size $|\lambda| = R$) and λ^* is the conjugate partition (or transposed Young diagram).

However, the classification of cohomological representations we just alluded to implies that very few of the irreducible submodules of $\wedge^\bullet \mathfrak{p}^*$ can occur as refined Hodge types of non-trivial cohomology classes. This is very analogous to the Kodaira vanishing theorem. The proof indeed makes a crucial use of a 'Dirac inequality' due to Parthasarathy, see Borel and Wallach [2000, Lemma II.6.11 and §II.7]. The vanishing theorem thus obtained generalizes a celebrated result of Matsushima [1962].

The K-types $V(\mathfrak{q})$ that can occur are determined by *admissible* pairs of partitions (λ, μ) i.e. partitions λ and μ as in (3-4) and such that if λ (resp. μ) is on the top left (resp. bottom right) corner of the rectangle $p \times q$ as pictured below (with $p = 4$, $q = 7$, $\lambda = (6, 6, 2, 0)$ and $\mu = (5, 2, 1, 0)$), the complementary boxes form a disjoint union of rectangles $p_1 \times q_1 \cup \ldots \cup p_r \times q_r$ (in the example below $1 \times 2 \cup 1 \times 3 \cup 1 \times 1$), see Bergeron [2009, Lemme 6].

•	•	•	•	•	•	
•	•	•	•	•	•	*
•	•				*	*
		*	*	*	*	*

We denote by $V(\lambda, \mu)$ the corresponding K-type. In particular the K-module $V(\lambda) := V(\lambda, 0)$ is isomorphic to $S_\lambda(V_+) \otimes S_{\lambda^*}(V_-)^*$. In general $V(\lambda, \mu)$ is isomorphic to the Cartan product of $V(\lambda)$ and $V(\mu)^*$. The first degree where such a K-type can occur in the cohomology is $R = |\lambda| + |\mu|$. More precisely, it contributes to the cohomology of bi-degree $(|\lambda|, |\mu|)$ in the Hodge-Lefschetz decomposition and we have:

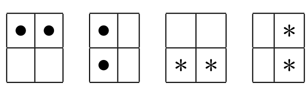
$$H^\bullet(\mathfrak{g}, K; A_\mathfrak{q}) \cong H^{\bullet - R}(\mathrm{Gr}_{p_1}(\mathbb{C}^{p_1+q_1}) \times \ldots \times \mathrm{Gr}_{p_r}(\mathbb{C}^{p_r+q_r}), \mathbb{C}).$$

Matsushima's formula (3-2) then strongly refines the Hodge-Lefschetz decomposition of compact quotients $\Gamma \backslash S$.

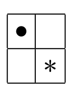
Example. Take $p = 2$ and $q = 2$. Compact quotients $\Gamma \backslash S$ are 4-dimensional complex manifolds. Their Betti numbers satisfy the relation $b_k = b_{8-k}$ because of Poincaré duality. They moreover decompose as sums $b_k = \sum_{p+q=k} h^{p,q}$ of Hodge numbers that satisfy $h^{p,q} = h^{q,p}$. But more is true: the vector $(b_0, \ldots, b_8)$ of Betti numbers of a compact quotient $\Gamma \backslash S$ is actually of the form

$$\begin{pmatrix}1\\0\\1\\0\\2\\0\\1\\0\\1\end{pmatrix} + 2h^{2,0}\begin{pmatrix}0\\0\\1\\0\\1\\0\\1\\0\\0\end{pmatrix} + (h^{1,1}-1)\begin{pmatrix}0\\0\\1\\0\\2\\0\\1\\0\\0\end{pmatrix} + 2(h^{3,0}+h^{2,1})\begin{pmatrix}0\\0\\0\\1\\0\\1\\0\\0\\0\end{pmatrix} + k\begin{pmatrix}0\\0\\0\\0\\1\\0\\0\\0\\0\end{pmatrix} \tag{3-5}$$

for some integer $k \geq 0$. The first vector indeed corresponds to the component of the trivial representation in Matsushima's formula. The second term corresponds to the components of the cohomological representations $A_\mathfrak{q}$ with $R = 2$ that contribute either to the holomorphic or anti-holomorphic cohomology. Their associated pairs of partitions (λ, μ) are



The third term corresponds to the components of the (unique) cohomological representation $A_\mathfrak{q}$ with $R = 2$ that contributes to the cohomology of bi-degree $(1, 1)$. Its associated pair of partitions (λ, μ) is



And so on...

3.3.2 Compact quotients of the symmetric space associated to $\mathrm{SO}_0(p, q)$**.** Even though these are not Hermitian in general, Matshushima's formula still makes sense. Considerations, similar to those in the unitary case, show that cohomological representations of G are essentially[1] parametrized by partitions $\lambda = (\lambda_1, \dots, \lambda_p)$, with $q \geq \lambda_1 \geq \dots \geq \lambda_p \geq 0$, such that the pair (λ, λ) is admissible.

Example. Take $p = 5$ and $q = 4$. Compact quotients $\Gamma \backslash S$ are 20-dimensional real manifolds. Their Betti numbers satisfy the relation $b_k = b_{20-k}$ because of Poincaré duality. But more is true: the Betti numbers of a compact quotient $\Gamma \backslash S$ actually verify the relations

$$b_1 = b_2 = b_3 = 0, \quad b_8 \geq 2b_6 \quad \text{and} \quad b_{10} \geq 3b_6. \tag{3-6}$$

Hyperbolic manifolds of dimension n correspond to $p = n$ and $q = 1$. Then Matsushima's formula essentially gives no restrictions on the Betti numbers.[2]

Among the family of symmetric spaces associated to $\mathrm{SO}_0(p, q)$, the ones where $q = 2$ are – up to exchanging the roles of p and q – the only Hermitian spaces; these are of complex dimension $n = p$. In these cases, $K \subset \mathrm{O}_n \times \mathrm{O}_2$ acts on $\mathfrak{p} = \mathbb{C}^n \otimes (\mathbb{C}^2)^*$ through the standard representation of O_n on $\mathbb{C}^n$ and the standard representation of O_2 on $\mathbb{C}^2$. Denote by $\mathbb{C}^+$ and $\mathbb{C}^-$ the $\mathbb{C}$-span of the vectors $e_1 + ie_2$ and $e_1 - ie_2$ in $\mathbb{C}^2$. The two lines $\mathbb{C}^+$ and $\mathbb{C}^-$ are left stable by O_2. This yields a decomposition $\mathfrak{p} = \mathfrak{p}^+ \oplus \mathfrak{p}^-$ which corresponds to the decomposition given by the natural complex structure on $\mathfrak{p}_0$. For each non-negative integer k the K-representation $\wedge^k \mathfrak{p} = \wedge^k(\mathfrak{p}^+ \oplus \mathfrak{p}^-)$ decomposes as the sum:

$$\wedge^k \mathfrak{p} = \bigoplus_{r+s=k} \wedge^r \mathfrak{p}^+ \otimes \wedge^s \mathfrak{p}^-.$$

The K-representations $\wedge^r \mathfrak{p}^+ \otimes \wedge^s \mathfrak{p}^-$ are not irreducible in general: there is at least a further splitting given by the Lefschetz decomposition:

$$\wedge^r \mathfrak{p}^+ \otimes \wedge^s \mathfrak{p}^- = \bigoplus_{\ell=0}^{\min(r,s)} \tau_{r-\ell, s-\ell}.$$

One can check that for $2(r+s) < n$ each K-representation $\tau_{r,s}$ is irreducible. Moreover in the range $2(r+s) < n$ only those with $r = s$ can occur as a K-type $V(\mathfrak{q})$ associated to a cohomological representation. One can moreover check that each $\tau_{r,r}$ is irreducible as long as $r < n$; it is isomorphic to some $V(\mathfrak{q})$ and corresponds to the partition $\lambda = (2_r, 0_{n-r})$.[3] Let us denote by $A_{r,r}$ the corresponding cohomological representation. We have:

$$H^{i,j}(\mathfrak{g}, K; A_{r,r}) = \begin{cases} \mathbb{C} & \text{if } r \leq i = j \leq n - r,\ 2i \neq n \\ \mathbb{C} + \mathbb{C} & \text{if } 2i = 2j = n \\ 0 & \text{otherwise.} \end{cases}$$

[1] This is completely true only if both p and q are odd.

[2] To be precise it gives no restriction at all if n is odd and one recovers that $b_{n/2}$ is even if n is even.

[3] When $2(r+s) \geq n$ the partitions $(2_r, 1_s, 0_{n-r-s})$ also correspond to cohomological representations.

In the particular case where n is even and $(r,r) = (0,0)$ – so that $A_{0,0}$ is the trivial representation – we have $H^\bullet(\mathfrak{g}, K; A_{0,0}) = H^\bullet(S^c, \mathbb{C})$, where $S^c = \mathrm{SO}(n+2)/\mathrm{SO}(n)\times \mathrm{SO}(2)$ is the complex quadric. The space $H^\bullet(S^c, \mathbb{C})$ has a basis $\{1, c_1, c_1^2, \dots, c_1^{n-1}, e\}$, where c_1 is the Chern class of the complexification of the line bundle arising from the standard representation of $\mathrm{SO}(2)$, i.e. the Kähler form on S^c, and where e is the Euler class of the vector bundle arising from the standard representation of $\mathrm{SO}(n)$.

4 Betti numbers of locally symmetric manifolds

One may wonder:

> what are the Betti numbers of a random locally symmetric space ?

A classical theorem of Gromov (see Ballmann, Gromov, and Schroeder [1985]) bounds from above the Betti numbers of a locally symmetric space by a constant (depending only of the dimension) times its volume. It is therefore natural to investigate the growth of the Betti numbers as the volume tends to infinity. The analogous question for complex hypersurfaces in $\mathbb{P}^{n+1}$ is classical.

4.1 Comparison with projective hypersurfaces. The fundamental projective invariant of an n-dimensional algebraic variety $V \subset \mathbb{P}^N$ is its degree d which is also equal to the volume – with respect to the standard Kähler form on $\mathbb{P}^N$ – divided by $n!$.

In case $V \subset \mathbb{P}^{n+1}$ is an hypersurface, by standard arguments involving Lefschetz Hyperplane Theorem and Poincaré duality (see e.g. Gayet and Welschinger [2014, Lemma 3]), we have $b_k(V) = b_k(\mathbb{P}^n)$ for $k \neq n$. On the other hand, the Euler-Poincaré characteristic of V is equal to

$$\chi(V) = \langle c_n(T_V), [V]\rangle = \frac{1}{d}\left[(1-d)^{n+2} - 1\right] + n + 2.$$

It follows that the growth of the Betti numbers of V with respect to the degree d is given by

$$b_k(V) = \begin{cases} O(1) & \text{if } k \neq n \\ (-1)^n \chi(V) + O(1) = d^{n+1} + O(d^n) & \text{if } k = n. \end{cases} \tag{4-1}$$

4.2 Asymptotics of Betti numbers of locally symmetric manifolds. It is not obvious at all that large volume locally symmetric S-manifolds should have related topological behavior. However, one consequence of Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017] and Abert, Bergeron, Biringer, and Gelander [n.d.] is the following theorem that is analogous to (4-1).

Theorem 1. *Suppose that G has property (T) and rank at least two. The growth of the Betti numbers of locally symmetric* S-manifolds *is given by*

$$b_k(\Gamma\backslash S) = \begin{cases} o\left(\mathrm{vol}(\Gamma\backslash S)\right) & \text{if } k \neq \frac{1}{2}\dim S \\ \frac{\chi(S^c)}{\mathrm{vol}(S^c)}\mathrm{vol}(\Gamma\backslash S) + o\left(\mathrm{vol}(\Gamma\backslash S)\right) & \text{if } k = \frac{1}{2}\dim S. \end{cases}$$

Example. Let $n \geq 3$ and let (Γ_m) be a sequence of distinct torsion-free lattices in $\mathrm{SL}_n(\mathbb{R})$. Then for all k, we have $b_k(\Gamma_m) = o(\mathrm{vol}(\Gamma_m\backslash \mathrm{SL}_n(\mathbb{R})))$ as $m \to +\infty$.

Hyperbolic spaces have a rank one group of isometries and it is not hard to construct examples of large volume hyperbolic manifolds with very different topologies, see e.g. Bergeron [2017] for many examples. This allows in particular to construct counter-examples to the conclusion of Theorem 1. However recent works of Fraczyk [2016] and Fraczyk and Raimbault [n.d.] imply that this conclusion holds for *congruence arithmetic* hyperbolic manifolds. More generally they prove:

Theorem 2. *Let S be arbitrary. The growth of the Betti numbers of* congruence arithmetic S-manifolds *is given by*

$$b_k(\Gamma\backslash S) = \begin{cases} o\left(\mathrm{vol}(\Gamma\backslash S)\right) & \text{if } k \neq \frac{1}{2}\dim S \\ \frac{\chi(S^c)}{\mathrm{vol}(S^c)}\mathrm{vol}(\Gamma\backslash S) + o\left(\mathrm{vol}(\Gamma\backslash S)\right) & \text{if } k = \frac{1}{2}\dim S. \end{cases}$$

Outside the middle degree it is hard to guess what should be the 'true' growth rate of the Betti numbers. For congruence arithmetic real hyperbolic manifolds $\Gamma\backslash\mathbb{H}^n$, associated to a fixed rational group $\mathbf{H}$ (see Section 2.2), it was suggested by Gromov (see Sarnak and Xue [1991]) that

$$b_k(\Gamma\backslash\mathbb{H}^n) \ll_{\mathbf{H},\varepsilon} \mathrm{vol}(\Gamma\backslash\mathbb{H}^n)^{\frac{2k}{n-1}+\varepsilon}. \tag{4-2}$$

Cossutta and Marshall [2013] suggest – and actually prove in a quite general situation – that the best exponent is in fact $2j/n$ as long as $k \neq (n \pm 1)/2$. See Marshall [2014] for similar results on other classes of symmetric spaces.

Remark. In a way that is quite similar to Bismut's proof of Demailly's asymptotic Morse inequalities (see Bismut [1987] and Demailly [1985]) for projective varieties, the existence of an upper bound sublinear in the volume is related (see e.g. the influential Sarnak and Xue [1991]) to the existence of a *spectral gap* for the Hodge-Laplace operator acting on differential k-forms with $k \neq (n \pm 1)/2$:

Theorem 3. *Let k be different from* $(n \pm 1)/2$. *There exists a positive constant* $\varepsilon = \varepsilon(n,k)$ *such that for any congruence arithmetic real hyperbolic manifolds* $\Gamma\backslash\mathbb{H}^n$*, the first non-zero eigenvalue of the Hodge-Laplace operator of* $\Gamma\backslash\mathbb{H}^n$ *acting on differentiable k-forms is bounded below by* ε.

This was conjectured in Bergeron and Clozel [2005] and proved in Bergeron and Clozel [2013].

When $k = (n \pm 1)/2$ there is no spectral gap – the corresponding cohomological representation of G is 'tempered' – and we do not know what to expect for the growth of $b_k(\Gamma\backslash\mathbb{H}^n)$. However for particular sequences of Γ's, Calegari and Emerton [2009] were able to prove an upper bound sublinear in the volume, see also Bergeron, Linnell, Lück, and Sauer [2014].

4.3 Explicit computations. Matsuhima's formula and the classification of cohomological representations imply many restrictions on the Betti numbers (e.g. in small degree some vanishing results or equality with the corresponding Betti numbers of S^c). Apart from these restrictions, explicit computations of the Betti numbers of a fixed locally symmetric space $\Gamma\backslash S$ in terms of the algebraic data defining Γ is usually a challenge. Very few cases are known. One of the first results of this type is the computation, by J.-S. Li [1996], of the dimension of the L^2-cohomology space of degree g of certain congruence arithmetic quotients of the Siegel upper half space of genus g.

The proof is divided into two parts. First, relying on previous works of Howe, Jian-Shu Li proves that the cohomology is generated by certain theta series. Then he computes the dimension of the space generated by these theta series. More recently in Bergeron, Millson, and Moeglin [2017] and Bergeron, Z. Li, Millson, and Moeglin [2017] we were able to prove that a large part of the cohomology of certain locally symmetric spaces associated to $\mathrm{SO}_0(n,2)$ is generated by certain theta series. Using previous computations by Bruinier [2002] of dimensions of the spaces generated by these theta series we get explicit expressions for certain Betti numbers. We prove in particular:

Theorem 4. *The rank of the Picard group of the moduli space $\mathcal{K}_g$, defined in Section 2.3, is*

$$\frac{31g+24}{24} - \frac{1}{4}\alpha_g - \frac{1}{6}\beta_g - \sum_{k=0}^{g-1}\left\{\frac{k^2}{4g-4}\right\} - \sharp\left\{k \mid \frac{k^2}{4g-4} \in \mathbb{Z}, 0 \le k \le g-1\right\}$$

where

$$\alpha_g = \begin{cases} 0, & \text{if } g \text{ is even,} \\ \left(\frac{2g-2}{2g-3}\right) & \text{otherwise,} \end{cases} \qquad \beta_g = \begin{cases} \left(\frac{g-1}{4g-5}\right) - 1, & \text{if } g \equiv 1 \mod 3, \\ \left(\frac{g-1}{4g-5}\right) + \left(\frac{g-1}{3}\right) & \text{otherwise,} \end{cases}$$

and $\left(\frac{a}{b}\right)$ is Jacobi symbol.

5 Cycle theory

Let M be a (closed) manifold. So far, we have computed the cohomology of M using smooth differential forms. We could as well have used currents. The resulting cohomology groups $H^k(M,\mathbb{C})$ are the same (and similarly for the groups occurring in the Hodge-Lefschetz or Matsushima decompositions). If Z is a closed orientable submanifold of real co-dimension k, it is an integral cycle and, by Poincaré duality, it defines a class $\mathrm{cl}(Z)$ in $H^k(M,\mathbb{C})$. The integration current on Z is closed of degree k and represents the image of $\mathrm{cl}(Z)$ in $H^k(M,\mathbb{C})$.

By a classical theorem of Thom, any class in the rational cohomology groups $H^k(M,\mathbb{Q})$ is a rational multiple of the cycle class $\mathrm{cl}(Z)$ of a (maybe disconnected) co-dimension k closed submanifold. When M is locally symmetric, it is natural to ask if one can restrict our choices of closed submanifold, e.g. to certain locally symmetric subspaces associated to subgroups $H \subset G$.

5.1 Comparison with projective varieties. In case M is a projective non-singular algebraic variety $V \subset \mathbb{P}^N$ over $\mathbb{C}$, it is natural to restrict to closed analytic subspaces $Z \subset V$, or equivalenly, by Chow's theorem, to algebraic cycles. Let p be the complex co-dimension of Z in V. Two analytic subvarieties of complementary dimension meeting in isolated points have a non-negative local intersection number. Since we can find a linear subspace $\mathbb{P}^{N-p}$ in $\mathbb{P}^N$ meeting V in isolated points, it follows that the cycle class $\mathrm{cl}(Z)$ is non-zero in $H^{2p}(V,\mathbb{C})$. Now the integration current on Z is closed of type (p,p). The class $\mathrm{cl}(Z)$ in $H^{2p}(V,\mathbb{C})$ is hence of type (p,p). Rational (p,p)-classes are called *Hodge classes*. They form the group $\mathrm{Hdg}^p(V,\mathbb{Q}) = H^{2p}(V,\mathbb{Q}) \cap H^{p,p}(V)$, and Hodge posed the famous:

Hodge Conjecture. On a projective non-singular algebraic variety over $\mathbb{C}$, any Hodge class is a rational linear combination of cycle classes $\mathrm{cl}(Z)$ of algebraic cycles.

Hodge also proposed a further conjecture, characterizing the subspace of $H^\bullet(V,\mathbb{Q})$ spanned by the images of cohomology classes with support in a suitable closed analytic subspace of complex codimension k. Grothendieck observed that this further conjecture is false, and gave a corrected version of it in Grothendieck [1969].

5.2 Modular and symmetric cycle classes. Let us come back to locally symmetric manifolds $\Gamma\backslash S$. To any connected center-free semi-simple closed subgroup $H \subset G$ corresponds an embedding of the symmetric space S_H associated to H into S. If $\Gamma \cap H$ is a lattice, the inclusion $S_H \hookrightarrow S$ induces an immersion of real analytic varieties $(\Gamma \cap H)\backslash S_H \to \Gamma\backslash S$ whose associated cycle class in $H^\bullet(\Gamma\backslash S,\mathbb{C})$ we denote by C_H^Γ. We will refer to these classes as *modular classes*. When $\Gamma\backslash S$ is non-compact, C_H^Γ is sometimes compactified to give a cycle on a natural compactification of $\Gamma\backslash S$ but we won't discuss these issues here.

Examples. 1. When S is the real hyperbolic n-space $\mathbb{H}^n$, the modular classes in $\Gamma\backslash\mathbb{H}^n$ are the cycle classes of totally geodesic immersed submanifold of finite volume.

2. In complex ball quotients, cycle classes of finite volume quotients of sub-balls give examples of modular classes. These are the only modular classes that are cycle classes of algebraic cycles but there might be other modular classes: to the inclusion $\mathrm{SO}_0(n,1) \subset \mathrm{PU}(n,1)$ corresponds a totally real geodesic embedding of the real hyperbolic n-space into the complex n-ball that may projects onto a non-zero modular classes.

3. The moduli space $\mathcal{K}_g$ of genus g quasi-polarized K3 surfaces – that identifies with a locally symmetric space associated to $G = \mathrm{SO}_0(2,19)$ – can have arbitrarily large Picard group (see Theorem 4) and, more generally, many classes of cycles in their Chow groups. In particular there are many cycles coming from Noether-Lefschetz theory: the locus parametrizing the K3 surfaces with Picard number strictly greater than some positive integer $r \leq 19 = \dim_{\mathbb{C}} \mathcal{K}_g$ is indeed a countable union of subvarieties of co-dimension r. The cycle classes of the irreducible components of this locus are modular classes associated to subgroups $H \subset G$ isomorphic to $\mathrm{SO}_0(2, 19-r)$. As in the case of ball quotients there are also non-algebraic modular classes.

There are a number of results on modular classes, but our current knowledge nevertheless appears to be quite poor: a large part of the literature on modular classes is only concerned in establishing the *non-vanishing* of these classes. As in the case of analytic subspaces of projective varieties, this has been addressed using the intersection numbers of these cycles, see e.g. Millson [1976], Millson and Raghunathan [1980], and Kudla and Millson [1990]. This has also been addressed using tools coming from representation theory, see especially Tong and Wang [1989] and J.-S. Li [1992]. This non-vanishing question is usually too hard to study for a given manifold; one simplifies the problem by 'stabilizing' it, that is to say by considering towers of finite coverings rather than a single manifold. Let us say that a modular class C_H^Γ is *virtually non-zero* if there exists a finite index subgroup $\Gamma' \subset \Gamma$ such that the modular class $C_H^{\Gamma'}$ is non-zero in $H^\bullet(\Gamma'\backslash S, \mathbb{C})$.

The following conjecture – see Bergeron [2006] for more details (in particular with respect to non-compact quotients $\Gamma\backslash S$) – provides a quite general answer to the question of the virtual non-vanishing of modular classes. To our knowledge this conjecture encompasses all known results. It has been (or can be) checked in most classical situations (see especially Bergeron [2006], Bergeron [2008], and Bergeron and Clozel [2013, 2017]). To formulate it, let us first distinguish some particular modular symbols. Say that a closed subgroup $H \subset G$ is a symmetric subgroup of G if there exists an involution τ of G such that $H = G^\tau$ is the connected component of the identity in the group of fixed points of τ. We will refer to the corresponding modular classes C_H^Γ as *symmetric modular classes*. All the modular classes from the examples above are symmetric.

Conjecture 5. *Assume for simplicity that* $\Gamma\backslash S$ *is compact. A symmetric modular class* C_H^Γ *is* virtually *non-zero in* the strongly primitive part of *the cohomology of degree* $\dim S - \dim S_H$ *if and only if* $\mathrm{rank}_{\mathbb{C}}(G/H) = \mathrm{rank}_{\mathbb{C}}(K/(K \cap H))$.

5.3 Hodge types of modular classes. Keeping in mind the analogy with projective varieties, the next step is to determine on which Hodge types of the cohomology of $\Gamma\backslash S$ modular classes can project non-trivially. Up to now it seems to have been addressed only in few particular cases. As explained in his 2002 ICM talk Kobayashi [2002], jointly with Oda, Kobayashi has devised a sufficient criterion for a modular class to be annihilated by a π-component in Matsushima's decomposition (3-2). Their proof is based on a theory of discrete branching laws for unitary representations of G. The most interesting cases they can deal with are compact quotients of the symmetric space $S = \mathrm{SO}_0(2n, 2)/\mathrm{SO}_{2n} \times \mathrm{SO}_2$. If $\Gamma\backslash S$ is a compact S-manifold, the contribution of the trivial representation of G to Matsushima's formula (3-2) yields a natural injective map of cohomology groups $H^\bullet(S^c, \mathbb{C}) \subset H^\bullet(\Gamma\backslash S, \mathbb{C})$; in particular we shall see the Euler class $e \in H^{n,n}(S^c, \mathbb{C})$, defined in Section 3.3.2, as an element in $H^{n,n}(\Gamma\backslash S, \mathbb{C})$. Kobayashi and Oda then prove:

Theorem 6. *The Hodge* (n, n)*-type component of a modular class* C_H^Γ *with* $H \cong \mathrm{SO}_0(2n, 1)$ *is proportional to the Euler class* e.

These modular classes are cycle classes of totally real, totally geodesic submanifolds of real dimension $2n$ into $\Gamma\backslash S$ which is a Kähler manifold of complex dimension $2n$. In case $n = 1$ the space S is a product $\mathbb{H}^2 \times \mathbb{H}^2$ and the cycles derived from H are obtained by 'partial complex conjugation' of algebraic cycles with respect to the complex conjugation on the second factor of S. Then Theorem 6 is equivalent to the well-known fact that the cycle class of a closed analytic (complex) co-dimension 1 subspace in a compact algebraic surface over $\mathbb{C}$ has no Hodge $(2, 0) + (0, 2)$-type components.

Beside the representation theoretic method of Kobayashi and Oda, a classical work of Kudla and Millson [1990] suggests another approach. Kudla and Millson indeed provide explicit dual forms to some natural modular classes in locally symmetric spaces associated to classical groups. From this, one can derive serious restrictions on the possible Hodge types to which these modular classes can contribute. Let's describe the two main families of examples.

5.3.1 Quotients of the symmetric space associated to $\mathrm{PU}(p, q)$**.** Let notations be as in Section 3.3.1. Let $c_q \in H^{q,q}(S^c, \mathbb{C})$ be the top Chern class of the q-dimensional vector bundle over $S^c = \mathrm{U}_{p+q}/\mathrm{U}_p \times \mathrm{U}_q$ associated to the standard representation of U_q, i.e. the q-th power of the Kähler form of S^c. Here again if $\Gamma\backslash S$ is a compact S-manifold, the contribution of the trivial representation of G to Matsushima's formula (3-2) yields a natural injective map of cohomology groups $H^\bullet(S^c, \mathbb{C}) \subset H^\bullet(\Gamma\backslash S, \mathbb{C})$ and we shall see the Chern class $c_q \in H^{q,q}(S^c, \mathbb{C})$ as an element in $H^{q,q}(\Gamma\backslash S, \mathbb{C})$. Wedging with c_q corresponds to applying the q-th power of the Lefschetz operator associated to the Kähler form on $\Gamma\backslash S$, and we define the subset $SH^\bullet(\Gamma\backslash S, \mathbb{C})$ of *special cohomology classes* in $H^\bullet(\Gamma\backslash S, \mathbb{C})$ by

$$SH^\bullet(\Gamma\backslash S, \mathbb{C}) = \bigoplus_{a,b=0}^{p} \bigoplus_{k=0}^{\min(p-a,p-b)} c_q^k H^{a\times q, b\times q}(\Gamma\backslash S, \mathbb{C}), \tag{5-1}$$

where $H^{a\times q, b\times q}(\Gamma\backslash S, \mathbb{C})$ denotes the generalized Hodge subspace of the cohomology corresponding to the pair of partitions (λ, μ) with λ an a by q rectangle and μ a b by q rectangle. As with the usual Hodge-Lefschetz decomposition, we have:

$$SH^\bullet(\Gamma\backslash S, \mathbb{C}) = \bigoplus_{a,b=0}^{p} SH^{aq,bq}(\Gamma\backslash S, \mathbb{C})$$

where the (usual) *primitive* part of the subspace $SH^{aq,bq}(\Gamma\backslash S, \mathbb{C})$ is exactly $H^{a\times q, b\times q}(\Gamma\backslash S, \mathbb{C})$.

Now the proof of Bergeron, Millson, and Moeglin [2016, Theorem 8.2] implies the following:

Proposition 7. *Let r be a non-negative integer with $r \leq p$ and let C_H^Γ be a modular class in $H^{2rq}(\Gamma\backslash S, \mathbb{C})$ with $H \cong \mathrm{PU}(p-r, q)$. Then C_H^Γ is an algebraic class and it is contained in* $\mathrm{SHdg}^r(\Gamma\backslash S, \mathbb{Q}) := SH^{rq,rq}(\Gamma\backslash S, \mathbb{C}) \cap H^{2rq}(\Gamma\backslash S, \mathbb{Q})$.

5.3.2 Quotients of the symmetric space associated to $\mathrm{SO}_0(p, q)$. Similarly and with notations as in Section 3.3.2, any modular class C_H^Γ, with H isomorphic to a smaller orthogonal group fixing a positive subspace, is contained in

$$SH^\bullet(\Gamma\backslash S, \mathbb{C}) = \oplus_{r=0}^{[p/2]} \oplus_{k=0}^{p-2r} e_q^k H^{r\times q}(\Gamma\backslash S, \mathbb{C}). \tag{5-2}$$

Here e_q is zero if q is odd and is the Euler class arising from the standard representation of SO_q if q is even. We then write $\mathrm{SHdg}^r(\Gamma\backslash S, \mathbb{Q}) = SH^{rq}(\Gamma\backslash S, \mathbb{C}) \cap H^{rq}(\Gamma\backslash S, \mathbb{Q})$.

Examples 1. If $q = 1$ the space S is the p-dimensional hyperbolic real space and the subspace $SH^\bullet(\Gamma\backslash S, \mathbb{C})$ is in fact equal to the full cohomology group $H^\bullet(\Gamma\backslash S, \mathbb{C})$.

2. If $q = 2$ the space is Hermitian and we have:

$$SH^\bullet(\Gamma\backslash S, \mathbb{C}) = \oplus_{r=0}^{p} H^{r,r}(\Gamma\backslash S, \mathbb{C}).$$

Beware that in this case the Euler class e_2 is the class of the Kähler form that we denoted c_1 in Section 3.3.2.

5.4 Hodge type theorems. Modular cycle classes belong to a subspace $\mathrm{SHdg}^\bullet(\Gamma\backslash S, \mathbb{Q})$ of the full cohomology group $H^\bullet(\Gamma\backslash S, \mathbb{C})$. Everything is therefore in place to raise a question analogous to the Hodge Conjecture:

> Do modular cycle classes span the subspace $\mathrm{SHdg}^\bullet(\Gamma\backslash S, \mathbb{Q})$?

We shall see that it is too much to hope for in general, but surprisingly enough this is close to be true in several interesting cases. Let us again consider our two main families of examples. Both cases are dealt with in joint works with Millson and Moeglin. The proofs rely heavily on Arthur's classification Arthur [2013] of automorphic representations of classical groups which depends on the stabilization of the trace formula for disconnected groups discussed in Waldspurger's 2014 ICM talk Waldspurger [2014] and recently obtained by Mœglin and Waldspurger [n.d.].

5.4.1 A Hodge type theorem for quotients of the symmetric space associated to $\mathrm{PU}(p,q)$. Even in the simple case where $p = 2$ and $q = 1$ – so that S is the complex 2-ball – it was proved by Blasius and Rogawski [2000] that there exist compact quotients $\Gamma\backslash S$ such that the space of Hodge $(1,1)$-classes is not spanned by modular classes. However, vaguely stated, the main result of Bergeron, Millson, and Moeglin [2016] asserts that for congruence arithmetic quotients $\Gamma\backslash S$ (with arbitrary p and q's) the special cohomology $SH^n(\Gamma\backslash S, \mathbb{C})$ is generated, for n small enough, by cup products of three types of classes:

- classes in $SH^{q,q}(\Gamma\backslash S, \mathbb{C})$;
- holomorphic and anti-holomorphic special cohomology classes, i.e. classes in $SH^{\bullet,0}(\Gamma\backslash S, \mathbb{C})$ and $SH^{0,\bullet}(\Gamma\backslash S, \mathbb{C})$;
- modular cycle classes of Section 5.3.1.

5.4.2 A Hodge type theorem for quotients of the symmetric space associated to $\mathrm{SO}_0(p,q)$. In that case, vaguely stated, the main result of Bergeron, Millson, and Moeglin [2017] states that as long as r is less than $\frac{1}{2}p$ and $\frac{1}{3}(p+q-1)$, the 'primitive' subspace $H^{r\times q}(\Gamma\backslash S, \mathbb{C})$ of $SH^{rq}(\Gamma\backslash S, \mathbb{C})$, in the decomposition (5-2), is spanned by projections of modular cycle classes.

5.5 Applications.

5.5.1 . The most striking consequences of the above mentioned 'Hodge type theorems' concern the cases where S is a complex ball or a real hyperbolic space. Indeed, in these cases $SH^\bullet(\Gamma\backslash S, \mathbb{C}) = H^\bullet(\Gamma\backslash S, \mathbb{C})$ and we obtain the two following theorems.

We first have to define more precisely the congruence arithmetic locally symmetric space we deal with: let E be either a totally real number field or a totally imaginary quadratic extension of a totally real number field. In both cases we denote by F the maximal totally real subfield of E. Now let V be a E-vector space of dimension $n+1 \geq 3$ and let $h : V \times V \to E$ be a Hermitian form with respect to the conjugaison of E/F, such that h is of signature $(n,1)$ at one real place of F and definite at the others. Let $\mathbf{H}$ be the semi-simple algebraic $\mathbb{Q}$-group obtained from the algebraic F-group $\mathrm{SU}(h)$ by restriction of scalars. To any congruence subgroup in $\mathbf{H}(\mathbb{Q})$ we attach a congruence arithmetic quotient $\Gamma\backslash S$ where S is the real hyperbolic n-space, if $E = F$, and the complex n-ball, otherwise.

Theorem 8. *Suppose that $\Gamma\backslash S$ is a closed complex n-ball quotient and let $r \in [0,n]\backslash]\frac{n}{3}, \frac{2n}{3}[$. Then every Hodge class in $H^{2r}(\Gamma\backslash S, \mathbb{Q})$ is algebraic.*

Remarks. 1. Beware that here modular cycle classes *do not* span, even in co-dimension 1. One has to consider arbitrary $(1,1)$-classes.

2. In small degree one can even confirm Hodge's generalized conjecture in its original formulation (with $\mathbb{Q}$ coefficients).

Theorem 9. *Suppose that $\Gamma\backslash S$ is a real hyperbolic n-manifold. Then, for all $r < n/3$, the $\mathbb{Q}$-vector space $H^r(\Gamma\backslash S, \mathbb{Q})$ is spanned by classes of totally geodesic cycles.*

Remarks. 1. In Bergeron, Millson, and Moeglin [2017] we provide strong evidence that Theorem 9 should not hold above the degree $n/3$.

2. When n is even, all congruence arithmetic real hyperbolic n-manifolds are of the simple type described above. However, when n is odd, there are other types of congruence arithmetic real hyperbolic n-manifolds. These *do not* contain totally geodesic immersed co-dimension 1 submanifolds. Still, they may have a non-zero first Betti number. Theorem 9 therefore cannot hold for general (congruence arithmetic) hyperbolic manifolds.

5.5.2 . When $G = \mathrm{SO}_0(n,2)$ the space S is Hermitian and our general 'Hodge type theorem' again specializes into new cases of the Hodge conjecture. Let us emphasize the even more special case of the moduli spaces $\mathcal{K}_g$ (in which cases we have $n = 19$): a theorem of Oguiso [2009] indeed implies that any curve on $\mathcal{K}_g$ meets some of the Noether-Lefschetz (NL) divisors described in Example (3) of Section 5.2. So it is natural to ask whether the Picard group $\mathrm{Pic}_{\mathbb{Q}}(\mathcal{K}_g)$ of $\mathcal{K}_g$ with rational coefficients is spanned by NL-divisors. This was conjectured to be true by Maulik and Pandharipande, see 'Noether-Lefschetz Conjecture' Maulik and Pandharipande [2013, Conjecture 3]. More generally, one can extend this question to higher NL-loci on $\mathcal{K}_g$. In Bergeron, Z. Li, Millson, and Moeglin [2017], we prove:

Theorem 10. *For all $g \geq 2$ and all $r \leq 4$, the cohomology group $H^{2r}(\mathcal{X}_g, \mathbb{Q})$ is spanned by NL-cyles of codimension r. In particular (taking $r = 1$),* $\mathrm{Pic}_{\mathbb{Q}}(\mathcal{X}_g) \cong H^2(\mathcal{X}_g, \mathbb{Q})$ *and the Noether-Lefschetz Conjecture holds on $\mathcal{X}_g$ for all $g \geq 2$.*

Remark. Before Bergeron, Millson, and Moeglin [2017], He and Hoffman [2012] considered another interesting special case of our general 'Hodge type theorem,' that of smooth Siegel modular threefolds Y where $p = 3$ and $q = 2$. They prove that $\mathrm{Pic}(Y) \otimes \mathbb{C} \cong H^{1,1}(Y)$ is generated by Humbert surfaces.

6 Automorphic Lefschetz properties

Almost 40 years ago Oda [1981] proved that the Albanese variety of a congruence arithmetic complex ball quotient is spanned by the Hecke translates of the Jacobian of a fixed Shimura curve. This implies a version of the Lefschetz Theorem on the injection of the cohomology to Shimura curves that can arguably be considered as the starting point of the analogy between locally symmetric spaces and projective varieties that we are discussing here. Since Oda's pioneering work, a number of criteria have been developed to determine if some Hecke translate of a given cohomology class on a locally symmetric space restricts non-trivially to a given locally symmetric subspace. Venkataramana's ICM 2010 talk Venkataramana [2010] was devoted to this subject. We shall therefore insist on results obtained since then.

6.1 Comparison with projective varieties. Let $V \subset \mathbb{P}^N$ be a projective non-singular algebraic variety and $V \cap H$ a hyperplane section of V. Then we have the

Lefschetz Hyperplane Theorem. *The restriction map*

(6-1)
$$H^i(V, \mathbb{Q}) \to H^i(V \cap H, \mathbb{Q})$$

is an isomorphism for $i \leq n - 2$ and injective for $i = n - 1$.

This theorem in fact contains two quite different statements:

1. the map (6-1) is injective for $i < n$,
2. the map $H_i(V \cap H, \mathbb{Q}) \to H_i(V, \mathbb{Q})$ is injective for $i \geq n$.

6.2 Restriction to special cycles. The Lefschetz Hyperplane Theorem applies to compact quotients $\Gamma \backslash S$ of *Hermitian* symmetric spaces but modular classes are not ample. However, one may consider all the translates of these modular classes under Hecke operators and ask for a weaker Lefschetz property for the collection of these Hecke translates.

And a large number of such ‘weak Lefschetz properties’ indeed hold even when the symmetric space S is not Hermitian, see e.g. Bergeron [2006]. Since the general results require a rather forbidding amount of notation we restrict our discussion to the special situation of Section 5.5.1.

In this situation, to any non-degenerate, indefinite, subspace $W \subset V$ defined over E we attach a special cycle $\Gamma_W \backslash S_W \to \Gamma \backslash S$. It is of real dimension $m[E : F]$. The following theorem is our analogue of the first part of Lefschetz Hyperplane Theorem.

Theorem 11. *For every $m < n$, there exist $(m+1)$-dimensional subspaces $W_1, \dots, W_s$ in V such that the restriction map*

$$H^i(\Gamma \backslash S, \mathbb{Q}) \to \bigoplus_j H^i(\Gamma_{W_j} \backslash S_{W_j}, \mathbb{Q}) \tag{6-2}$$

is injective for all $i < \frac{1}{2} m[E : F]$ and for $i = \frac{1}{2} m[E : F]$ if $\Gamma \backslash S$ is closed.

The theorem can be reformulated using Hecke correspondences: to any F-rational element g of the isometry group of h we may associate a finite correspondence $(\Gamma \cap g^{-1} \Gamma g) \backslash S \rightrightarrows \Gamma \backslash S$ where the first projection is the covering projection and the second projection is induced by the multiplication by g. Write $C_g^* : H^\bullet(\Gamma \backslash S, \mathbb{Q}) \to H^\bullet(\Gamma \backslash S, \mathbb{Q})$ for the induced endomorphism. Theorem 11 then says that if α is a non-zero class in $H^i(\Gamma \backslash S, \mathbb{Q})$ of degree $i < \frac{1}{2} \dim_{\mathbb{R}} S_W$, then there exists a g such that $C_g^*(\alpha)$ pulls back non-trivially to $\Gamma_W \backslash S_W$.

For compact ball quotients, the theorem is due to Oda [1981] in degree $i = 1$, and to Venkataramana [2001] – confirming a conjecture of Harris and J.-S. Li [1998] – for all degrees. The essential point (in the case $n = m + 1$) is that a linear combination of the divisors $\Gamma_{W_i} \backslash S_{W_i} \to \Gamma \backslash S$ gives a particular ample class, the hyperplane class in the canonical projective embedding of $\Gamma \backslash S$. The Lefschetz property then follows from the hard Lefschetz theorem. For non-compact ball quotients, one can combine Venkataramana’s idea with the study of compactifications, see Nair [2017]. It may appear quite surprising that the theorem holds for *real* hyperbolic manifolds. This is again a topological consequence of the spectral gap Theorem 3. This approach gives a unified proof of Theorem 11, see Bergeron and Clozel [2013, 2017].

6.3 Another type of Lefschetz property.

As for the second part of Lefschetz Hyperplane Theorem, let us mention the following homotopical analogue of it, see Bergeron, Haglund, and Wise [2011].

Theorem 12. *A closed arithmetic hyperbolic manifold $\Gamma \backslash \mathbb{H}^n$ virtually retracts onto any of its* co-dimension 1 *modular cycle.*

In other words: if $\Lambda\backslash\mathbb{H}^{n-1} \to \Gamma\backslash\mathbb{H}^n$ is a totally geodesic immersion and if we write $\iota : \Lambda \to \Gamma$ for the corresponding (injective) morphism, there exists a finite index subgroup $\Gamma' \subset \Gamma$ and a morphism $r : \Gamma' \to \Lambda$ such that $\iota(\Lambda)$ is contained in Γ' and $r \circ \iota : \Lambda \to \Lambda$ is the identity map. In particular the induced map

$$H_i(\Lambda\backslash\mathbb{H}^{n-1}, \mathbb{Q}) \to H_i(\Gamma'\backslash\mathbb{H}^n, \mathbb{Q})$$

is injective for all $i \geq 0$.

The proof of Theorem 12 is very specific to arithmetic hyperbolic manifolds that contain co-dimension 1 modular cycles; it uses that the group Γ may be 'cubulated' (in the sense of Wise [2014]). It is a very interesting open question to decide which lattices of $SO_0(n, 1)$ can be cubulated, but it is known that lattices in all other real simple Lie groups cannot. Of course, this does not prevent Theorem 12 to hold for other locally symmetric spaces, but to my knowledge no other examples are known to (homotopically) retract onto a locally symmetric proper subspace except for a small finite number of beautiful examples, due to Deraux [2011], of complex 2-ball and 3-ball quotients that retract onto one of their totally geodesic submanifolds.

However, thanks to spectral gap properties as in Theorem 3, the homological consequences of Theorem 12 are more tractable in general, see Bergeron [2006]. In the special situation of Section 5.5.1 one can for example prove the following analogue of the second aspect of Lefschetz Hyperplane Theorem, see Bergeron and Clozel [2013].

Theorem 13. *Suppose that $\Gamma\backslash S$ is closed. Let W be a subspace of V. There exists a finite index subgroup $\Gamma' \subset \Gamma$ such that the natural map*

$$H_i(\Gamma'_W\backslash S_W, \mathbb{Q}) \to H_i(\Gamma'\backslash S, \mathbb{Q}) \tag{6-3}$$

is injective for all $i \geq \frac{1}{2}\dim_{\mathbb{R}} S$.

6.4 Some refined analogies with specific projective varieties: Abelian varieties. Theorem 9 is an analogue, in constant negative curvature, of the classical fact that cycle classes of totally geodesic flat sub-tori span the cohomology groups of flat tori. In fact if A is an Abelian variety, in most interesting cohomology theories $H^\bullet(A)$ is an exterior algebra on $H^1(A)$. In particular, if A is sufficiently general, the algebra of Hodge classes is generated in degree 1 and the Hodge conjecture follows.

Quite surprisingly, in small degrees, the cohomology rings of congruence arithmetic locally symmetric manifolds enjoy structural properties very analogous to those of Abelian varieties. Here again we discuss only the special situation of Section 5.5.1. Suppose furthermore that $\Gamma\backslash S$ is closed and write $\mathrm{Hdg}^\bullet(\Gamma\backslash S, \mathbb{Q}) = H^\bullet(\Gamma\backslash S, \mathbb{Q})$ if $E = F$, i.e. if S is a real hyperbolic space $\mathbb{H}^n$. First, the proofs of Theorems 8 and 9 imply:

Theorem 14. *The natural morphism of algebras*

$$\wedge^\bullet \mathrm{Hdg}^1(\Gamma\backslash S, \mathbb{Q}) \to \mathrm{Hdg}^\bullet(\Gamma\backslash S, \mathbb{Q}) \tag{6-4}$$

is onto in degree $< n/3$.

As opposed to what happens with Abelian varieties, the map (6-4) is not injective in general (already when $\Gamma\backslash S$ is a real hyperbolic surface). The next theorem – see Bergeron and Clozel [2013, 2017] – nevertheless shows that it is injective 'up to Hecke correspondences.'

Theorem 15. *Let* α *and* β *two cohomology classes in* $H^\bullet(\Gamma\backslash S, \mathbb{Q})$ *of respective degrees* k *and* ℓ *with* $k+\ell \leq \frac{1}{2}\dim_{\mathbb{R}} S$. *Then, there exists some rational element* g *of the isometry group of* h *such that*

$$C_g^*(\alpha) \wedge \beta \neq 0 \text{ in } H^{k+\ell}(M, \mathbb{Q}).$$

For complex ball quotients Theorem 15 is due to Venkataramana [2001]. Parthasarathy [1982], Clozel [1992, 1993] and Venkataramana [2010] have general results of this type for other Hermitian spaces, see also Bergeron [2004]. In Bergeron [2006] we consider more general non Hermitian locally symmetric spaces. Here again the key input is a spectral gap theorem.

7 Periods

Algebraic varieties admit a panoply of cohomology theories, related over $\mathbb{C}$ by comparison isomorphisms. These give rise to different structures on the cohomology groups. Comparing two such structures leads in particular to the rich theory of periods.

When dealing with general locally symmetric manifolds we don't have all these cohomology theories at our disposal anymore. However, using the canonical Riemannian structure on S, we can extract some numerical invariants from the cohomology, which we call 'period matrices.'

7.1 Comparison with projective varieties. If V is a smooth projective variety defined over $\mathbb{Q}$, the vector space $H^k(V, \mathbb{C})$ has a natural $\mathbb{Q}$-structure $H^k_{\mathrm{dR}}(X/\mathbb{Q})$: choose a cover of V by Zariski affine open sets defined over $\mathbb{Q}$ and use algebraic differential forms with coefficients in $\mathbb{Q}$. A comparison theorem, due to Grothendieck, gives a natural isomorphism $H^k_{\mathrm{dR}}(X/\mathbb{Q}) \otimes \mathbb{C} \cong H^k(V, \mathbb{C})$.

One calls *periods* the matrix coefficients of the comparison isomorphism

$$H^k_{\mathrm{dR}}(X/\mathbb{Q}) \otimes \mathbb{C} \xrightarrow{\cong} H^k(V, \mathbb{Q}) \otimes \mathbb{C}$$

between algebraic de Rham cohomology and singular cohomology after choosing $\mathbb{Q}$-bases in both groups. In general these two different $\mathbb{Q}$-structures are transcendent with respect to each other and periods are fundamental numerical invariants, see e.g. Kontsevich and Zagier [2001].

7.2 Period matrices of locally symmetric spaces. Let us come back to locally symmetric manifolds $\Gamma\backslash S$. Through the isomorphism (3-1), the Riemannian structure on S induces a positive definite quadratic form on each cohomology group $H^j(\Gamma\backslash S, \mathbb{Z})$ modulo torsion. Letting $b = b_j(\Gamma\backslash S)$, we encode the above data into a matrix

$$M = \left(\int_{\gamma_k} \omega_\ell \right)_{1 \le k, \ell \le b} \in \mathrm{GL}_b(\mathbb{R})$$

where the $\gamma_k \in H_j(\Gamma\backslash S, \mathbb{Z})$ project to a basis for $H_j(\Gamma\backslash S, \mathbb{Z})$ modulo torsion and the ω_ℓ's are an orthonormal basis for the space of harmonic j-forms on $\Gamma\backslash S$. The matrix M is well-defined up to multiplication on the left by $\mathrm{GL}_b(\mathbb{Z})$ and on the right by an orthogonal matrix.

As an element of $\mathrm{GL}_b(\mathbb{Z})\backslash\mathrm{GL}_b(\mathbb{R})/\mathrm{O}_b$, the matrix M is characterized by its determinant and its image in the locally symmetric space that parametrizes the space of flat b-dimensional tori of unit volume. In analogy with the classical Schottky problem, it would be interesting to analyse the locus of the latter as Γ varies. Let us restrict our attention to the apparently simpler question of bounding the determinant.

7.3 Regulators. Following Bergeron and Venkatesh [2013] and Bergeron, Şengün, and Venkatesh [2016] we call 'degree j regulator' the determinant of the degree j period matrix of $\Gamma\backslash S$; we denote it by $R_j(\Gamma\backslash S)$.

Note that $|R_0(\Gamma\backslash S)| = 1/\sqrt{\mathrm{vol}(\Gamma\backslash S)}$, $|R_n(\Gamma\backslash S)| = \sqrt{\mathrm{vol}(\Gamma\backslash S)}$, and by Poincaré duality, we have $|R_j(\Gamma\backslash S) R_{n-j}(\Gamma\backslash S)| = 1$. We propose the following:

Conjecture 16. *Fix S and j. The growth of the degree j regulators of* congruence arithmetic *of S-manifolds is given by*

$$\log |R_j(\Gamma\backslash S)| = o(\mathrm{vol}(\Gamma\backslash S)).$$

In the next paragraph we relate Conjecture 16 to the geometric complexity of cycles needed to generate $H_j(\Gamma\backslash S, \mathbb{R})$. 'Hodge type theorems' like Theorem 9 suggest that the conjecture could be tractable when j is far enough from the middle degree. In general one can think of Conjecture 16 as an attempt to shed little light on the mysterious cycle theory of locally symmetric spaces near the middle degree.

7.4 Back to cycles. Our reason to believe in Conjecture 16 is that, roughly speaking, we expect homology classes on congruence arithmetic manifolds to be represented by cycles of low complexity. In our general situation, these cycles are not algebraic at all but one may still hope that their *topological* complexity reflects the *arithmetic* complexity of their (Langlands-)associated varieties.

In Bergeron, Şengün, and Venkatesh [2016] we formulate and study the following precise conjecture in a simple interesting case, namely, that of congruence arithmetic hyperbolic 3-manifolds.

Conjecture 17. *There is an absolute constant C such that, for any congruence arithmetic hyperbolic 3-manifold $\Gamma\backslash\mathbb{H}^3$, there exist immersed surfaces S_i of genus less that $\mathrm{vol}(\Gamma\backslash\mathbb{H}^3)^C$ such that the $[S_i]$'s span $H_2(\Gamma\backslash\mathbb{H}^3,\mathbb{R})$.*

To relate Conjecture 17 with $R_2(\Gamma\backslash\mathbb{H}^3)$, we study the relationship between two norms on the second homology group: the purely topological Gromov-Thurston norm and the more geometric 'harmonic' norm. Refining Bergeron, Şengün, and Venkatesh [ibid., Proposition 4.1] Brock and Dunfield [2017] show that these two norms are roughly proportional with explicit constants depending only on the volume and injectivity radius[4] of $\Gamma\backslash\mathbb{H}^3$. Now, assuming Conjecture 17, each $[S_j]$ has Gromov-Thurston norm – and therefore harmonic norm – which is bounded by a polynomial in $\mathrm{vol}(\Gamma\backslash\mathbb{H}^3)$. Thus Hadamard's inequality shows that $|R_2(\Gamma\backslash\mathbb{H}^3)| \ll \mathrm{vol}(\Gamma\backslash\mathbb{H}^3)^{Cb_1(\Gamma\backslash\mathbb{H}^3)}$.

Remarks. 1. Conjectures 16 and 17 are false if the manifolds are not assumed to be congruence arithmetic: Brock and Dunfield [ibid., Theorem 1.5] indeed construct a sequence of closed hyperbolic 3-manifolds M_n (whose injectivity radii stay bouded away from 0) with

$$\mathrm{vol}(M_n) \to \infty,\ b_1(M_n) = 1 \text{ and } \limsup_n \frac{\log|R_2(M_n)|}{\mathrm{vol}(M_n)} > 0.$$

2. Under well believed number theoretic assumptions, in Bergeron, Şengün, and Venkatesh [2016] we notably verify Conjecture 17 when $\Gamma\backslash\mathbb{H}^3$ is a congruence cover of a Bianchi manifold with 1-dimensional cuspidal cohomology associated to a non-CM elliptic curve. In that case the proof indeed relate the complexity of the H_2-cycle to the height of the associated elliptic curve (i.e., the minimal size of A, B so that its equation can be expressed as $y^2 = x^3 + Ax + B$). That this might be a general phenomenon was also suggested in Calegari and Venkatesh [2012].

[4] which is expected to be uniformly bounded away from 0 on arithmetic manifolds

8 Some regrets

Many other interesting questions could (should?) have been discussed in this survey. Hyperbolic 3-manifolds form a particularly rich and interesting family of locally symmetric manifolds. However: Matsushima's formula gives no restriction on their cohomology and most of these manifolds do not contain totally geodesic immersed submanifolds. It may therefore appear that hyperbolic 3-manifolds are not really connected with our general story. This is not quite true: Agol [2013, 2014] proof of the celebrated 'Virtual Haken Conjecture' suggests considering 'almost geodesic' cycles rather than just geodesic ones. We haven't addressed the rich relation between the cohomology of locally symmetric spaces and number theory. Let us simply say that our original motivation for Conjectures 16 and 17 came from the study of torsion homology and its relation with Galois representations Scholze [2014]. Finally Venkatesh's program Venkatesh [2017] suggests fascinating relations between the period matrices of Section 7.2 and periods (in the usual sense) of automorphic forms.

Acknowledgments. I would like to thank Miklos Abert, Tsachik Gelander, Laurent Clozel, Étienne Ghys, Frédéric Haglund, Zhiyuan Li, John Millson, Colette Moeglin, Peter Sarnak, T.N. Venkataramana, Akshay Venkatesh, Claire Voisin and Dani Wise for many very helpful conversations over the years on the material related to this survey.

References

Miklos Abert, Nicolas Bergeron, Ian Biringer, and Tsachik Gelander (n.d.). "Non-uniformly discrete IRS and normallized Betti convergence". To appear (cit. on p. 858).

Miklos Abert, Nicolas Bergeron, Ian Biringer, Tsachik Gelander, Nikolay Nikolov, Jean Raimbault, and Iddo Samet (2017). "On the growth of L^2-invariants for sequences of lattices in Lie groups". *Ann. of Math. (2)* 185.3, pp. 711–790. MR: 3664810 (cit. on p. 858).

Ian Agol (2013). "The virtual Haken conjecture". *Doc. Math.* 18. With an appendix by Agol, Daniel Groves, and Jason Manning, pp. 1045–1087. MR: 3104553 (cit. on p. 873).

– (2014). "Virtual properties of 3-manifolds". In: *Proceedings of the International Congress of Mathematicians, Seoul, Vol. I*. Kyung Moon SA Co. Ltd., pp. 141–170 (cit. on p. 873).

James Arthur (2013). *The endoscopic classification of representations*. Vol. 61. American Mathematical Society Colloquium Publications. Orthogonal and symplectic groups. American Mathematical Society, Providence, RI, pp. xviii+590. MR: 3135650 (cit. on p. 865).

Werner Ballmann, Mikhael Gromov, and Viktor Schroeder (1985). *Manifolds of nonpositive curvature*. Vol. 61. Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, pp. vi+263. MR: 823981 (cit. on p. 858).

N. Bergeron (2008). "Représentations cohomologiques isolées, applications cohomologiques". *J. Inst. Math. Jussieu* 7.2, pp. 205–246. MR: 2400721 (cit. on p. 862).

– (2009). "Restriction de la cohomologie d'une variété de Shimura à ses sous-variétés". *Transform. Groups* 14.1, pp. 41–86. MR: 2480852 (cit. on p. 855).

Nicolas Bergeron (2004). "Produits dans la cohomologie des variétés de Shimura: quelques calculs". *C. R. Math. Acad. Sci. Paris* 339.11, pp. 751–756. MR: 2110376 (cit. on p. 870).

– (2006). "Propriétés de Lefschetz automorphes pour les groupes unitaires et orthogonaux". *Mém. Soc. Math. Fr. (N.S.)* 106, pp. vi+125. MR: 2296755 (cit. on pp. 862, 868–870).

– (2017). "Variétés en expansion (d'après M. Gromov, L. Guth, . . .)". *Séminaire Bourbaki*. 69 ème année, Exp. No. 1132, vii–viii, 115–150 (cit. on p. 859).

Nicolas Bergeron and Laurent Clozel (2005). "Spectre automorphe des variétés hyperboliques et applications topologiques". *Astérisque* 303, pp. xx+218. MR: 2245761 (cit. on p. 860).

– (2013). "Quelques conséquences des travaux d'Arthur pour le spectre et la topologie des variétés hyperboliques". *Invent. Math.* 192.3, pp. 505–532. MR: 3049928 (cit. on pp. 860, 862, 868–870).

– (2017). "Sur le spectre et la topologie des variétés hyperboliques de congruence: les cas complexe et quaternionien". *Math. Ann.* 368.3-4, pp. 1333–1358. MR: 3673656 (cit. on pp. 862, 868, 870).

Nicolas Bergeron, Frédéric Haglund, and Daniel T. Wise (2011). "Hyperplane sections in arithmetic hyperbolic manifolds". *J. Lond. Math. Soc. (2)* 83.2, pp. 431–448. MR: 2776645 (cit. on p. 868).

Nicolas Bergeron, Zhiyuan Li, John J. Millson, and Colette Moeglin (2017). "The Noether–Lefschetz conjecture and generalizations". *Invent. Math.* 208.2, pp. 501–552. MR: 3639598 (cit. on pp. 860, 866).

Nicolas Bergeron, Peter Linnell, Wolfgang Lück, and Roman Sauer (2014). "On the growth of Betti numbers in p-adic analytic towers". *Groups Geom. Dyn.* 8.2, pp. 311–329. MR: 3231217 (cit. on p. 860).

Nicolas Bergeron, John J. Millson, and Colette Moeglin (2016). "The Hodge conjecture and arithmetic quotients of complex balls". *Acta Math.* 216.1, pp. 1–125. MR: 3508219 (cit. on pp. 864, 865).

– (2017). "Hodge Type Theorems for Arithmetic Manifolds Associated to Orthogonal Groups". *Int. Math. Res. Not. IMRN* 15, pp. 4495–4624. MR: 3685109 (cit. on pp. 860, 865–867).

Nicolas Bergeron, Mehmet Haluk Şengün, and Akshay Venkatesh (2016). "Torsion homology growth and cycle complexity of arithmetic manifolds". *Duke Math. J.* 165.9, pp. 1629–1693. MR: 3513571 (cit. on pp. 871, 872).

Nicolas Bergeron and Akshay Venkatesh (2013). "The asymptotic growth of torsion homology for arithmetic groups". *J. Inst. Math. Jussieu* 12.2, pp. 391–447. MR: 3028790 (cit. on p. 871).

Jean-Michel Bismut (1987). "Demailly's asymptotic Morse inequalities: a heat equation proof". *J. Funct. Anal.* 72.2, pp. 263–278. MR: 886814 (cit. on p. 859).

Don Blasius and Jonathan Rogawski (2000). "Cohomology of congruence subgroups of $\mathrm{SU}(2,1)^p$ and Hodge cycles on some special complex hyperbolic surfaces". In: *Regulators in analysis, geometry and number theory*. Vol. 171. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 1–15. MR: 1724885 (cit. on p. 865).

A. Borel and N. Wallach (2000). *Continuous cohomology, discrete subgroups, and representations of reductive groups*. Second. Vol. 67. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. xviii+260. MR: 1721403 (cit. on p. 855).

Jeffrey F. Brock and Nathan M. Dunfield (2017). "Norms on the cohomology of hyperbolic 3-manifolds". *Invent. Math.* 210.2, pp. 531–558. MR: 3714511 (cit. on p. 872).

Jan Hendrik Bruinier (2002). "On the rank of Picard groups of modular varieties attached to orthogonal groups". *Compositio Math.* 133.1, pp. 49–63. MR: 1918289 (cit. on p. 860).

F. Calegari and A. Venkatesh (Dec. 2012). "A torsion Jacquet–Langlands correspondence". arXiv: 1212.3847 (cit. on p. 872).

Frank Calegari and Matthew Emerton (2009). "Bounds for multiplicities of unitary representations of cohomological type in spaces of cusp forms". *Ann. of Math. (2)* 170.3, pp. 1437–1446. MR: 2600878 (cit. on p. 860).

Donald I. Cartwright and Tim Steger (2010). "Enumeration of the 50 fake projective planes". *C. R. Math. Acad. Sci. Paris* 348.1-2, pp. 11–13. MR: 2586735 (cit. on p. 852).

Shiing-shen Chern (1957). "On a generalization of Kähler geometry". In: *Algebraic geometry and topology. A symposium in honor of S. Lefschetz*. Princeton University Press, Princeton, N. J., pp. 103–121. MR: 0087172 (cit. on p. 849).

L. Clozel (1992). "Produits dans la cohomologie holomorphe des variétés de Shimura". *J. Reine Angew. Math.* 430, pp. 69–83. MR: 1172908 (cit. on p. 870).

– (1993). "Produits dans la cohomologie holomorphe des variétés de Shimura. II. Calculs et applications". *J. Reine Angew. Math.* 444, pp. 1–15. MR: 1241791 (cit. on p. 870).

Mathieu Cossutta and Simon Marshall (2013). "Theta lifting and cohomology growth in p-adic towers". *Int. Math. Res. Not. IMRN* 11, pp. 2601–2623. MR: 3065089 (cit. on p. 859).

P. Deligne and G. D. Mostow (1986). "Monodromy of hypergeometric functions and non-lattice integral monodromy". *Inst. Hautes Études Sci. Publ. Math.* 63, pp. 5–89. MR: 849651 (cit. on p. 852).

Jean-Pierre Demailly (1985). "Champs magnétiques et inégalités de Morse pour la d''-cohomologie". *Ann. Inst. Fourier (Grenoble)* 35.4, pp. 189–229. MR: 812325 (cit. on p. 859).

Martin Deraux (2011). "Forgetful maps between Deligne-Mostow ball quotients". *Geom. Dedicata* 150, pp. 377–389. MR: 2753711 (cit. on p. 869).

M. Fraczyk (Dec. 2016). "Strong Limit Multiplicity for arithmetic hyperbolic surfaces and 3-manifolds". arXiv: 1612.05354 (cit. on p. 859).

M. Fraczyk and J. Raimbault (n.d.). "Convergence of arithmetic locally symmetric spaces". To appear (cit. on p. 859).

William Fulton (1997). *Young tableaux*. Vol. 35. London Mathematical Society Student Texts. With applications to representation theory and geometry. Cambridge University Press, Cambridge, pp. x+260. MR: 1464693 (cit. on p. 855).

Damien Gayet and Jean-Yves Welschinger (2014). "What is the total Betti number of a random real hypersurface?" *J. Reine Angew. Math.* 689, pp. 137–168. MR: 3187930 (cit. on p. 858).

A. Grothendieck (1969). "Hodge's general conjecture is false for trivial reasons". *Topology* 8, pp. 299–303. MR: 0252404 (cit. on p. 861).

Michael Harris (2014). "Automorphic Galois representations and the cohomology of Shimura varieties". In: *Proceedings of the International Congress of Mathematicians, Seoul, Vol. II*. Kyung Moon SA Co. Ltd., pp. 367–389 (cit. on p. 850).

Michael Harris and Jian-Shu Li (1998). "A Lefschetz property for subvarieties of Shimura varieties". *J. Algebraic Geom.* 7.1, pp. 77–122. MR: 1620690 (cit. on p. 868).

Hongyu He and Jerome William Hoffman (2012). "Picard groups of Siegel modular 3-folds and θ-liftings". *J. Lie Theory* 22.3, pp. 769–801. MR: 3012154 (cit. on p. 867).

Bruno Klingler (2003). "Sur la rigidité de certains groupes fondamentaux, l'arithméticité des réseaux hyperboliques complexes, et les "faux plans projectifs"". *Invent. Math.* 153.1, pp. 105–143. MR: 1990668 (cit. on p. 852).

Toshiyuki Kobayashi (2002). "Branching problems of unitary representations". In: *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*. Higher Ed. Press, Beijing, pp. 615–627. MR: 1957069 (cit. on p. 863).

Maxim Kontsevich and Don Zagier (2001). "Periods". In: *Mathematics unlimited—2001 and beyond*. Springer, Berlin, pp. 771–808. MR: 1852188 (cit. on p. 871).

Stephen S. Kudla and John J. Millson (1990). "Intersection numbers of cycles on locally symmetric spaces and Fourier coefficients of holomorphic modular forms in several complex variables". *Inst. Hautes Études Sci. Publ. Math.* 71, pp. 121–172. MR: 1079646 (cit. on pp. 862, 863).

Jian-Shu Li (1992). "Nonvanishing theorems for the cohomology of certain arithmetic quotients". *J. Reine Angew. Math.* 428, pp. 177–217. MR: 1166512 (cit. on p. 862).

– (1996). "On the dimensions of spaces of Siegel modular forms of weight one". *Geom. Funct. Anal.* 6.3, pp. 512–555. MR: 1392328 (cit. on p. 860).

Simon Marshall (2014). "Endoscopy and cohomology growth on $U(3)$". *Compos. Math.* 150.6, pp. 903–910. MR: 3223876 (cit. on p. 859).

Yozô Matsushima (1962). "On Betti numbers of compact, locally sysmmetric Riemannian manifolds". *Osaka Math. J.* 14, pp. 1–20. MR: 0141138 (cit. on p. 855).

Davesh Maulik and Rahul Pandharipande (2013). "Gromov-Witten theory and Noether-Lefschetz theory". In: *A celebration of algebraic geometry*. Vol. 18. Clay Math. Proc. Amer. Math. Soc., Providence, RI, pp. 469–507. MR: 3114953 (cit. on p. 866).

John J. Millson (1976). "On the first Betti number of a constant negatively curved manifold". *Ann. of Math. (2)* 104.2, pp. 235–247. MR: 0422501 (cit. on p. 862).

John J. Millson and M. S. Raghunathan (1980). "Geometric construction of cohomology for arithmetic groups. I". In: *Geometry and analysis*. Indian Acad. Sci., Bangalore, pp. 103–123. MR: 592256 (cit. on p. 862).

C. Mœglin and J.-L. Waldspurger (n.d.). *Stabilisation de la formule des traces tordue*. available at https://webusers.imj-prg.fr/~colette.moeglin/ (cit. on p. 865).

Arvind N. Nair (2017). "Lefschetz properties for noncompact arithmetic ball quotients II". *Manuscripta Math.* 152.3-4, pp. 443–457. MR: 3608300 (cit. on p. 868).

Takayuki Oda (1981). "A note on the Albanese variety of an arithmetic quotient of the complex hyperball". *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* 28.3, 481–486 (1982). MR: 656032 (cit. on pp. 867, 868).

Keiji Oguiso (2009). "Picard number of the generic fiber of an abelian fibered hyperkähler manifold". *Math. Ann.* 344.4, pp. 929–937. MR: 2507632 (cit. on p. 866).

R. Parthasarathy (1982). "Holomorphic forms in $\Gamma\backslash G/K$ and Chern classes". *Topology* 21.2, pp. 157–178. MR: 641998 (cit. on p. 870).

Émile Picard (1881). "Sur une extension aux fonctions de deux variables du problème de Riemann relatif aux fonctions hypergéométriques". *Ann. Sci. École Norm. Sup. (2)* 10, pp. 305–322. MR: 1508705 (cit. on p. 852).

Gopal Prasad and Sai-Kee Yeung (2009). "Arithmetic fake projective spaces and arithmetic fake Grassmannians". *Amer. J. Math.* 131.2, pp. 379–407. MR: 2503987 (cit. on p. 852).

Peter Sarnak and Xiao Xi Xue (1991). "Bounds for multiplicities of automorphic representations". *Duke Math. J.* 64.1, pp. 207–227. MR: 1131400 (cit. on p. 859).

Peter Scholze (2014). "Perfectoid spaces and their applications". In: *Proceedings of the International Congress of Mathematicians, Seoul, Vol. II*. Kyung Moon SA Co. Ltd., pp. 461–486 (cit. on p. 873).

Birgit Speh (2006). “Representation theory and the cohomology of arithmetic groups”. In: *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, pp. 1327–1335. MR: 2275647 (cit. on p. 850).

William P. Thurston (1998). “Shapes of polyhedra and triangulations of the sphere”. In: *The Epstein birthday schrift*. Vol. 1. Geom. Topol. Monogr. Geom. Topol. Publ., Coventry, pp. 511–549. MR: 1668340 (cit. on p. 852).

Y. L. Tong and S. P. Wang (1989). “Geometric realization of discrete series for semisimple symmetric spaces”. *Invent. Math.* 96.2, pp. 425–458. MR: 989703 (cit. on p. 862).

T. N. Venkataramana (2001). “Cohomology of compact locally symmetric spaces”. *Compositio Math.* 125.2, pp. 221–253. MR: 1815394 (cit. on pp. 868, 870).

– (2010). “Cohomology of arithmetic groups and representations”. In: *Proceedings of the International Congress of Mathematicians. Volume III*. Hindustan Book Agency, New Delhi, pp. 1366–1375. MR: 2827845 (cit. on pp. 850, 867, 870).

Akshay Venkatesh (2017). “Cohomology of arithmetic groups and periods of automorphic forms”. *Jpn. J. Math.* 12.1, pp. 1–32. MR: 3619577 (cit. on pp. 850, 873).

David A. Vogan Jr. and Gregg J. Zuckerman (1984). “Unitary representations with nonzero cohomology”. *Compositio Math.* 53.1, pp. 51–90. MR: 762307 (cit. on p. 854).

J.-L. Waldspurger (2014). “Stabilisation de la partie géométrique de la formule des traces tordue”. In: *Proceedings of the International Congress of Mathematicians, Seoul, Vol. II*. Kyung Moon SA Co. Ltd., pp. 487–504 (cit. on p. 865).

Daniel T. Wise (2014). “The cubical route to understanding groups”. In: *Proceedings of the International Congress of Mathematicians, Seoul, Vol. II*. Kyung Moon SA Co. Ltd., pp. 1075–1099 (cit. on p. 869).

Received 2017-11-12.

Nicolas Bergeron
Sorbonne Université
Institut de Mathématiques de Jussieu–Paris Rive Gauche
CNRS, Univ Paris Diderot
Paris
nicolas.bergeron@imj-prg.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (877–902)

COMPLEX BRUNN–MINKOWSKI THEORY AND POSITIVITY OF VECTOR BUNDLES

Bo Berndtsson

Abstract

This is a survey of results on positivity of vector bundles, inspired by the Brunn-Minkowski and Prékopa theorems. Applications to complex analysis, Kähler geometry and algebraic geometry are also discussed.

1 Introduction

The classical Brunn-Minkowski theorem is an inequality for the volumes of convex sets. It can be formulated as a statement about how the volumes of the vertical n-dimensional sections of a convex body in $\mathbb{R}^{n+m}$ vary with the section; more precisely it says that the n:th root of the volumes define a concave function on $\mathbb{R}^m$. The theorem has many applications, and it has also been generalized in many different directions (see e. g. the survey, Gardner [2002]).

One important generalization is Prékopa's theorem, Prékopa [1973], which can be seen as a version of the Brunn-Minkowski theorem for convex functions instead of convex sets. Let $\phi(x,t)$ be a convex function on $\mathbb{R}^n_x \times \mathbb{R}^m_t$, satisfying some mild extra conditions. We define a new function on $\mathbb{R}^m$ by

$$e^{-\tilde{\phi}(t)} = \int_{\mathbb{R}^n} e^{-\phi(t,x)} dx.$$

Then Prékopas theorem says that $\tilde{\phi}$ is also convex. Measures of the type

$$e^{-\phi} dx,$$

where ϕ is convex, are called *log concave* , and in this terminology Prékopa's theorem says that the marginals, or push forwards, of log concave measures are log concave. If we

Partially supported by grants from Vetenskapsrådet.
MSC2010: primary 32L05; secondary 14D06.

admit convex functions that attain the value ∞, the Brunn-Minkowski theorem is a direct consequence of Prékopa's theorem, corresponding to the case when ϕ is the indicator function of a convex body, i.e. the function which is zero on the convex body and infinity outside. (Properly speaking we get a version of the Brunn-Minkowski theorem with n:th roots replaced by logarithms. This is the 'multiplicative version' of the Brunn-Minkowski theorem, and is easily seen to be equivalent to the 'additive version'.)

Among the many different proofs of Prékopa's theorem, the one that is relevant to us here is the proof of Brascamp and Lieb [1976]. It goes by first proving the *Brascamp-Lieb inequality*, which is a Poincaré inequality for the d-operator (on functions). In the case when $n = 1$ (which is really the main case) it says the following: Let u be a smooth function on $\mathbb{R}$, satisfying

$$\int_{\mathbb{R}} u e^{-\phi(x)} dx = 0,$$

where ϕ is a smooth function, which is strictly convex in the sense that $\phi'' > 0$. Then

$$\int_{\mathbb{R}} |u|^2 e^{-\phi} dx \leq \int_{\mathbb{R}} \frac{|u'|^2}{\phi''} e^{-\phi} dx.$$

From this, Prékopa's theorem (and therefore also the Brunn-Minkowski theorem) follows just by differentiating $\tilde{\phi}$ twice, and applying the Brascamp-Lieb inequality to estimate the result. Actually, the two results are 'equivalent' in the sense that the Brascamp-Lieb theorem follows from Prékopa's theorem too. (See Cordero-Erausquin and Klartag [2012], which also gives a nice complement to the oversimplified historical picture described above.)

Now we observe that the Brascamp-Lieb inequality is the real variable version of Hörmander's L^2-estimate for the $\bar\partial$-operator, Hörmander [1965], Demailly [1997]. (This point of view was probably first stated clearly in Cordero-Erausquin [2005].) To describe the simplest case of Hörmander's estimate we let ϕ be a smooth strictly subharmonic function in $\mathbb{C}$. Then we let u be a smooth function on $\mathbb{C}$, satisfying

$$\int_{\mathbb{C}} u \bar{h} e^{-\phi} d\lambda = 0,$$

for all holomorphic functions h satisfying the appropriate L^2-condition ($d\lambda$ denotes Lebesgue measure). Note that this is a direct counterpart to the condition on u in the real case. Then u was assumed to be orthogonal to all constant functions, i.e. all functions in the kernel of d, whereas in the complex case u is assumed to be orthogonal to all functions in the kernel of $\bar\partial$. The conclusion of Hörmander's estimate is now that

$$\int_{\mathbb{C}} |u|^2 e^{-\phi} d\lambda \leq \int_{\mathbb{C}} \frac{|\bar\partial u|^2}{\Delta\phi} d\lambda,$$

which clearly is very similar to the conclusion of the Brascamp-Lieb inequality. The condition that u is orthogonal to all holomorphic functions, means that u is the L^2-minimal solution to a $\bar\partial$-equation, and this is how Hörmander's theorem is mostly thought of; as an estimate for solutions to the $\bar\partial$-equation. (There is also an even more important existence part of Hörmander's theorem, but that plays no role here.) In the same way, the Brascamp-Lieb theorem is an L^2-estimate for the d-equation, and it can be obtained as a special case of Hörmander's theorem, when the functions involved do not depend on the imaginary part of z.

Given the importance of Prékopa's theorem in convex geometry, it now becomes a natural question if there are any analogous consequences of Hörmander's theorem in the complex setting, that generalize the real theory. This is the subject of the work that we will now describe.

The most naive generalization would be that letting ϕ be plurisubharmonic (i.e. subharmonic on each complex line) in $\mathbb{C}^n \times \mathbb{C}^m$, and defining $\tilde\phi$ as we did in the real case, we get a plurisubharmonic function. This is however in general not the case as shown by a pertinent example of Kiselman [1978],

$$\phi((t,z) := |z-\bar t|^2 - |t|^2 = |z|^2 - 2\mathrm{Re}\, zt$$

(in $\mathbb{C}^2$). It turns out that instead we should think of the volume of a (convex) domain as the squared L^2-norm of the function 1, and the integrals of $e^{-\phi}$ as squared weigthed L^2-norms. Then it becomes natural to consider L^2-norms of holomorphic functions in the complex case, i.e. to look at norms

$$\|h\|_t^2 := \int_{\mathbb{C}^n} |h(z)|^2 e^{-\phi(t,z)} d\lambda(z),$$

or similar expressions where we integrate over slices of pseudoconvex domains in $\mathbb{C}^n$ instead of the total space. Let A_t^2 be the (Bergman) space of holomorphic functions with finite norm. Then we get a family of Hilbert spaces, indexed by t, i.e. a vector bundle, or at least a 'vector bundle like' object (the bundles obtained are in general not locally trivial). The theorems that we are going to discuss amount to saying that the curvature of these bundles is positive, or at least non negative, under natural assumptions. If, intuitively, we think of the curvature as (the negative of) the Hessian of the (logarithm of) the metric, this can be seen as a counterpart to the Brunn-Minkowski-Prékopa theorem. One can also recover Prékopa's theorem as a special case, see Section 2.

Let us add one remark on the relation of positive curvature to convexity in the real setting. A convex function on $\mathbb{R}^n$ is the same thing as a plurisubharmonic function on $\mathbb{C}^n$ that does not depend on the imaginary part of z. But, it is *not* the case that

$$-\log \|h_t\|_t^2,$$

is plurisubharmonic in t if h_t is a holomorphic section of a holomorphic hermitian vector bundle of positive curvature. This does however hold if the rank of the vector bundle is 1, so that we have a line bundle, if we also assume that $h_t \neq 0$ for all t. This is precisely the situation in the real setting: The bundle of constants has rank 1 and the 'section' 1 is never zero, and that is why we get simpler statements in the real setting. If we imagine a vector valued Brunn-Minkowski theory (see Raufi [n.d.]) we would get a situation similar to the complex case since

$$-\log \sum_1^N e^{-\phi_j}$$

is in general not convex for convex ϕ_j, except when $N = 1$.

Looking at the complex situation, it is clear that the restriction to linear sections of a domain in Euclidean space is not as natural as in the real case. The general picture involves two complex manifolds X and Y of dimensions $n+m$ and m respectively, and a holomorphic surjective map $p : X \to Y$ between them. This corresponds to the previous situation when $X = \mathbb{C}^n \times \mathbb{C}^m$, $Y = U \subset \mathbb{C}^m$ and p is the linear projection from $\mathbb{C}^n \times \mathbb{C}^m$ to $\mathbb{C}^m$. The role of the linear sections is played by the fibers $X_y = p^{-1}(y)$ of the map. To get enough holomorphic objects to apply the theory to, we will also replace holomorphic functions by holomorphic sections of a line bundle, L, over X, and the plurisubharmonic function ϕ now corresponds to a hermitian metric of non negative curvature on L. We then have almost all the ingredients for the complex theory, but one item remains to sort out: How do we define L^2-norms over the fibers?

For a general complex manifold, like the fibers of the map p, there is of course no substitute for Lebesgue measure. The way out is to consider, instead of sections of L, $(n,0)$-forms on the fibers with values in L. Such forms with values in L have natural L^2-norms, defined by wedge product and the metric on L. The bundle metric we get is

$$\|u\|_y^2 = c_n \int_{X_y} u \wedge \bar{u} e^{-\phi}.$$

Again, in our model situation of Euclidean space, this corresponds to integration with respect to Lebesgue measure. With this list of translations we obtain a counterpart to Prékopa's theorem in the complex setting, under natural convexity assumptions on X.

As it turns out, the nicest situation is when the map is proper, so that the fibers are compact, and also a submersion, so that the fibers are manifolds. In this case, and assuming also that the line bundle is trivial, the theorem was already known. Indeed it is a part of Griffiths' monumental theory of variations of Hodge structures, Griffiths and Tu [1984], Fujita [1978]. Griffiths' point of view however was rather different. He considered the vector bundle that we are discussing, with fibers $H^{n,0}(X_y)$, as a subbundle of the Hodge bundle with fibers $H^n_{dRh}(X_y)$, with connection induced by the Gauss-Manin connection on the Hodge bundle. It seems difficult to generalize this approach completely to the twisted case, when L is nontrivial, since e.g. there is no twisted counterpart to the Hodge bundle (see however Kawamata [1998]). It is also interesting to notice the different roles played by holomorphic forms: Griffiths was interested in holomorphic forms per se because of their relation to the period map. Here the forms are forced upon us in order to define L^2-norms.

In the next section we will give more precise formulations of the basic results. After that we will turn to applications; to L^2-extension problems for holomorphic functions, the (Mabuchi) space of Kähler metrics in a fixed cohomology class, variations of complex strucures, and finally positivity of direct image bundles in algebraic geometry.

Acknowledgments. The work described here has been influenced by discussions with many people. In particular I owe a great debt to Robert Berman, Sebastien Boucksom, Dario Cordero-Erausquin, Bo'az Klartag, Mihai Paun, Yanir Rubinstein and Xu Wang.

2 The basic results

We begin with non proper maps, and start by looking at the simplest case. We let Ω be a domain in $\mathbb{C}^n$ and U a domain in $\mathbb{C}^m$. We then let $X = \Omega \times U$ and let $p : X \to U$ be the linear projection on the second factor. We will assume that Ω is pseudoconvex (meaning that it has a smooth strictly plurisubharmonic exhaustion function). Let $\phi = \phi(z,t)$ be plurisubharmonic and smooth up to the boundary. Let

$$A^2(\Omega, \phi) = \{u \in H(\Omega); \int_\Omega |u|^2 e^{-\phi(\cdot,t)} < \infty\},$$

be the corresponding Bergman space of holomorphic functions, equipped with the norm

$$\|u\|_t^2 = \int_\Omega |u|^2 e^{-\phi(\cdot,t)}.$$

Since ϕ is bounded, all the Bergman spaces are the same as linear spaces, but their norms vary with t. We therefore get a trivial vector bundle $E = A^2 \times U$ over U with a non trivial Hermitian metric. We define its complex structure by saying that a section $t \to h(t, \cdot)$ is holomorphic if h is holomorphic on $\Omega \times U$.

Theorem 2.1. *(Berndtsson [2009]) The bundle $(E, \|\cdot\|_t)$ has non negative (Chern) curvature in the sense of Nakano.*

There are two main notions of positivity for Hermitian vector bundles; positivity in the sense of Griffiths and in the sense of Nakano. Positivity in the sense of Griffiths is defined in terms of the curvature tensor in the following way. Recall that the curvature is a $(1,1)$-form, Θ, with values in the space of endomorphisms of E. Then, if $u_t \in E_t$,

$$\langle \Theta u_t, u_t \rangle_t$$

is a scalar valued $(1,1)$-form. Positivity in the sense of Griffiths means that it is a positive form for any u_t. For the definition of positivity in the sense of Nakano, which is a stronger notion, we refer to Demailly [1997] or Berndtsson [2009]. We should also point out that we are here in the somewhat non standard situation of infinite rank bundles; this is also discussed in Berndtsson [ibid.].

We first discuss the proof very briefly. Let F be the bundle $(L^2(\Omega,\phi), \|\cdot\|_t)$, whose fibers are general L^2-functions, not necessarily holomorphic. It is also a trivial bundle, and we define a section $t \to u_t$ to be holomorphic if the dependence on t is holomorphic. Then E is a subbundle of F. It is easy to see that the curvature of F is the $(1,1)$-form $\sum_1^m \phi_{t_j,\bar{t}_k} dt_j \wedge d\bar{t}_k$, where the coefficients should be interpreted as the endomorphisms

$$u_t \to \phi_{t_j,\bar{t}_k} u_t.$$

Then it is immediately clear that F has non negative curvature as soon as ϕ is plurisubharmonic in t for z fixed. By general principles (Demailly [1997]), the curvature of the subbundle E is given by

$$\langle \Theta^E u_t, u_t \rangle_t = \langle \Theta^F u_t, u_t \rangle_t - \|\pi_{E^\perp} \theta^F u_t\|_t^2.$$

Here θ^F is the connection form for the Chern connection of F, and $\pi_{E^\perp}$ is the orthogonal projection on the orthogonal complement of E in F. The important thing to notice is now that, since $\pi_{E^\perp}\theta^F u_t$ lies in the orthogonal complement of the space of holomorphic functions, it is the L^2-minimal solution of some $\bar\partial$-equation, and we can apply Hörmander's estimate (for pseudoconvex domains in $\mathbb{C}^n$). This allows us to control the second, negative, term by the first, positive, one, and that gives the theorem.

It is well known that positivity in the sense of Griffiths, is equivalent to negativity of the dual bundle, E^*. On the other hand, negativity in the sense of Griffiths, is equivalent to saying that

$$\log \|v_t\|_t$$

is plurisubharmonic for any holomorphic section of E^*.

This leads to a more concrete reformulation of the first theorem. For t in U, let $t \to \mu_t$ be a family of complex measures on Ω. Assume there is a compact subset, K, of Ω, such that all μ_t are supported on K. Then

$$u_t \to \int_\Omega u_t d\mu_t =: \mu_t(u_t)$$

defines a secion of the dual bundle E^* and it gets a norm inherited from the norm on E_t.

Corollary 2.2. *Assume that the section μ_t is holomorphic in the sense that*

$$t \to \mu_t(h(t, \cdot)$$

is holomorphic for any h holomorphic on X. Then

$$\log \|\mu_t\|_t$$

is plurisubharmonic.

But this statement makes sense in a much more general situation.

Theorem 2.3. *Let D be a pseudoconvex domain in $\mathbb{C}^n \times \mathbb{C}^m$, and denote by $D_t = \{z \in \mathbb{C}^n; (z, t) \in D\}$ its vertical slices. Let μ_t be a family of measures on D_t that are all locally supported in a compact subset of D, and depend on t in a holomorphic way so that*

$$t \to \mu_t(h(t, \cdot))$$

is holomorphic in t if h is holomorphic in D. Then

$$\log \|\mu_t\|_t$$

is plurisubharmonic in t.

This way we have implicitly defined positivity of curvature for the 'bundle of Bergman spaces' $A^2(D_t, \phi)$, even though this is not properly speaking a vector bundle, since it is not locally trivial. Here we also note that Theorem 2.3 and Corollary 2.2 imply Prékopa's theorem: Take $D = (\mathbb{C}^*)^n$ (where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$), and define μ by taking averages over the n-dimensional real torus

$$\mu(h) = \int_{T^n} h(e^{i\theta_1} z_1, ..e^{i\theta_n} z_n) d\theta$$

(it does not depend on z). Computing the norm of μ as a functional on $A^2(D, e^{-\phi(t,\cdot)})$, where ϕ only depends on $|z_j|$, we recover Prékopa's theorem.

One main case of Theorem 2.3 is when the measures μ_t are all Dirac delta functions.

Theorem 2.4. *(Berndtsson [2006]) Let D be a pseudoconvex domain in $\mathbb{C}^n \times \mathbb{C}^m$ and let ϕ be plurisubharmonic in D. For each t in the projection of D to $\mathbb{C}^m$ let $B_t(z) = B_t(z, z)$ be the diagonal Bergman kernel for $A^2(D_t, \phi)$. Then $\log B_t(z)$ is plurisubharmonic in D.*

This theorem was obtained earlier when $n = 1$ and $\phi = 0$ by Maitani and Yamaguchi in Maitani and Yamaguchi [2004]. It is a consequence of the previous result. Let $t \to f(t)$

be a holomorphic map such that $f(t) \in D_t$, and let μ_t be a Dirac mass at $f(t)$. It is immediate that $t \to \mu_t(h) = h(t, f(t))$ is holomorphic if h is holomorphic on the product space. Moreover,

$$\|\mu_t\|_t^2 = B_t(f(t)),$$

so by the theorem, $\log B_t(f(t))$ is plurisubharmonic for any such map. This means that $B_t(z)$ is plurisubharmonic in D.

The proof of Theorem 2.3 uses the corollary, applied to a sequence of weights that tend to infinity outside of D. It is given in detail in Berndtsson [2006] for the situation of Bergman kernels, but the same proof gives the more general result too. See also Wang [2017] for a detailed analysis of a curvature operator defined in the setting of Theorem 2.3 (under one extra assumption) and conditions for when the curvature vanishes.

We next turn to general surjective holomorphic maps between complex manifolds, and there we will restrict to the case of proper maps, to fix ideas. Let X be a complex manifold of dimension $n+m$ and U a complex manifold of dimension m. Since our results are mostly local we can here think of U as a domain in $\mathbb{C}^m$. Let $p : X \to U$ be a holomorphic and surjective map. We say that p is smooth if its differential is surjective at every point. Then the fibers $X_t = p^{-1}(t)$ are smooth manifolds, and when the map is proper they are also compact. Let $L \to X$ be a holomorphic line bundle over X, equipped with a an Hermitian metric $e^{-\phi}$.

For t in the base, U, we let

$$E_t = H^{n,0}(X_t, L),$$

be the space of holomorphic L-valued $(n, 0)$-forms on X_t. We can also think of this space as

$$E_t = H^0(X_t, K_{X_t} + L),$$

the space of holomorphic sections over X_t of the canonical bundle of X_t twisted with the bundle L. Equivalently

$$E_t = H^0(X_t, K_{X/U} + L),$$

is the space of holomorphic sections over X_t of the relative canonical bundle $K_{X/U}$ twisted with L. (The relative canonical bundle, defined as $K_{X/U} = K_X - K_U$, is a line bundle over the total space X that restricts to K_{X_t} on every fiber X_t.) Let

$$E = \cup_{t \in U}\{t\} \times E_t.$$

It turns out that, when the metric on L has semipositive curvature, so that $i\partial\bar{\partial}\phi \geq 0$, then E is a holomorphic vector bundle over U. A section u_t of E is defined to be holomorphic if

$$u_t \wedge dt$$

is a holomorphic $(n+m,0)$ -form on X, where $(t_1, ...t_m)$ are holomorphic coordinates on U and $dt = dt_1 \wedge ...dt_m$. Finally, the L^2-norm

$$\|u_t\|_t^2 := c_n \int_{X_t} u_t \wedge \bar{u}_t e^{-\phi},$$

defines an Hermitian structure on E. We are then in a situation analogous to the one in Theorem 2.1, with the advantage that we now get a bona fide vector bundle of finite rank.

Theorem 2.5. *(Berndtsson [2009]) Assume that the curvature of L is nonnegative ($i\partial\bar{\partial}\phi \geq 0$) and that the total space is Kähler. Then E is a holomorphic vector bundle with non-negative curvature in the sense of Nakano (and hence also in the sense of Griffiths).*

Notice that what corresponds to pseudoconvexity here is the assumption that X is Kähler, so somewhat curiously the Kähler assumption appears as a convexity condition. If we look at the global picture and assume that X is even projective this is rather natural, since we may then remove a divisor from X and get a total space that is Stein, hence pseudoconvex. Since a divisor is removable for L^2-holomorphic forms, it appears that projectivity is related to (pseudo)convexity for these problems, but as it turns out, Kähler is enough.

When the fibration is trivial, so that $X = Z \times U$, the theorem can be proved in much the same way as Theorem 2.1, at least when the curvature of $e^{-\phi}$ is strictly positive on fibers. The proof in the general case is based on the formalism of pushforwards of currents. We write the norm squared of a holomorphic section as the pushforward of a form on the total space under p, compute $i\partial\bar{\partial}\|u_t\|_t^2$ using this formalism, and can then identify the curvature.

Theorem 2.5 can be developed in two different ways, by either adding more, or assuming less, assumptions. Let us first assume that the curvature form of the metric on L, $i\partial\bar{\partial}\phi$ restricts to a strictly positive form on each fiber. Then we have a Kähler metric $\omega_t := i\partial\bar{\partial}\phi|_{X_t}$ on each fiber, and thus an associated Laplace operator, $\Box_t = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$ on L-valued forms on each fiber. We can then give an 'explicit' formula for the curvature. To formulate the result we also assume that the base dimension $m = 1$; the general case follows by restricting to lines or curves in the base. We then define a $(1,1)$-form

$$C(\phi) = c(\phi) i dt \wedge d\bar{t},$$

by

$$(i\partial\bar{\partial}\phi)^{n+1}/(n+1)! = C(\phi) \wedge (i\partial\bar{\partial}\phi)^n/n!$$

(we will come back to this form later in sections 4 and 5).

Theorem 2.6. *(Berndtsson [2011]) Assume that $\omega_t = i\partial\bar{\partial}\phi|X_t > 0$ on each fiber X_t. Then the curvature, Θ of $(E, \|\cdot\|_t)$ is given by*

$$\langle \Theta_{\partial/\partial t, \partial/\partial \bar{t}} u_t, u_t \rangle_t = c_n \int_{X_t} c(\phi) u_t \wedge \bar{u}_t e^{-\phi} + \langle (1+\Box_t)^{-1} \kappa_\phi \cup u_t, \kappa_\phi \cup u_t \rangle_t$$

(where we define $\kappa_\phi \cup$ below).

To define κ_ϕ, we first have to define the notion of 'horizontal lift' of a vector field, $\partial/\partial t$ on the base, as introduced by Schumacher [2012], generalizing the earlier notion of 'harmonic lift' of Siu [1986]. A vector field on the total space X is said to be vertical, if it maps to zero under p. It is horizontal if it is orthogonal to all vertical fields under the (possibly degenerate, possible indefinite) metric $i\partial\bar\partial\phi$. As shown in Schumacher [2012], any vector field on the base U has a unique horizontal lift. In our case, the vector field on the base is $\partial/\partial t$, and there is a unique horisontal field V_ϕ (depending on $i\partial\bar\partial\phi$) which maps to $\partial/\partial t$ under p. Taking $\bar\partial V_\phi$ and restricting to fibers we get κ_ϕ; it is a $\bar\partial$-closed $(0,1)$-form on each fiber, with values in $T^{1,0}(X_t)$. The cohomology class of κ_ϕ in $H^{0,1}(X_t, T^{1,0}(X_t))$ is the Kodaira-Spencer class, but here it is important to look at this particular representative of the class. The cup product $\kappa_\phi \cup u_t$ is obtained by contracting with the vector part of κ_ϕ and wedging with the form part.

From the formula we see that if $i\partial\bar\partial\phi \geq 0$ and the curvature is zero, then $c(\phi) = 0$ and κ_ϕ is zero on each fiber. These two conditions imply that the horizontal lift of $\partial/\partial t$ is a holomorphic field, whose flow maps fibers to fibers. In fact, we have

Corollary 2.7. *Assume that $i\partial\bar\partial\phi \geq 0$, $i\partial\bar\partial\phi > 0$ on fibers and that the curvature Θ vanishes. Then there is a holomorphic vector field V on X whose flow lifts to L, such that its flow maps fibers of p to fibers and is an isometry on L. In particular, the flow maps ω_t to $\omega_{t'}$.*

In general terms, this means that the curvature can only vanish if the fibration $p : X \to U$ is trivial, holomorphically and metrically. This is proved under the assumption that the curvature of L is strictly positive on each fiber, and it does not hold in general without this assumtion (see Berndtsson [2011]).

It is also interesting to compare again to Prékopa's theorem. There, if the function $\tilde\phi$ is not strictly convex but linear (and if we assume that $m = 1$), then ϕ must have the form

$$\phi(t,x) = \psi(x + tv) + ct,$$

where ψ is a convex function on $\mathbb{R}^n$, v is a vector in $\mathbb{R}^n$ and c is a constant, Dubuc [1977]. Thus, the variation of ϕ with respect to t comes from the flow of a constant vector field applied to a fixed function ψ, 'lifted' to the line bundle $\mathbb{R}^n \times \mathbb{R}$ by adding ct. This is similar to what happens in the complex setting, except that there we get a holomorphic vector field instead.

We next discuss versions of Theorem 2.5 in the more general setting when the metric on the line bundle L is allowed to be singular, and the map p is no longer assumed to be smooth (in the sense described above) but only surjective. On the other hand, we now assume that X is projective and that $p : X \to Y$, where Y is a compact manifold.

By Sard's theorem, p is smooth outside of $p^{-1}(Y_1)$, where Y_1 is a proper analytic subvariety of Y, and there we can define the L^2-metric as before. There is a counterpart of the Bergman kernel here. It is first defined for $a \in p^{-1}(Y_1)$ as the norm of the evaluation functional at a on $E_{p(a)}$,

$$B(a) = \sup |u(a)|^2/\|u\|^2_{p(a)},$$

where the supremum is taken over all sections u in $H^0(X_{p(a)}, K_{X/Y} + L)$ that are square integrable with respect to the metric $e^{-\phi}$ (if there are no nontrivial such section, we let $B(a) = 0$). This definition however depends on the trivialization of $K_{X/Y} + L$ chosen near a, so the Bergman kernel is not a function but defines a metric on $K_{X/Y} + L$ – the Bergman kernel metric $B^{-1} = e^{-\log B}$.

Theorem 2.8. *(Berndtsson and Păun [2008])If the singular metric $e^{-\phi}$ has semipositive curvature (i.e. $i\partial\bar\partial\phi \geq 0$ in the sense of currents), and the Bergman kernel metric is not identically equal to 0, it defines a singular metric of semipositive curvature on $p^{-1}(Y_1)$. Moreover, this metric extends to a singular metric on $K_{X/Y} + L$ over all of X.*

Briefly, if L is pseudoeffective, i.e. has a singular metric of nonnegative curvature, $K_{X/Y} + L$ is also pseudoeffective, provided that it has a non trivial L^2-section over at least one fiber.

This result can be seen as a counterpart to Theorem 2.3 in this setting. The more difficult problem of counterparts of Theorem 2.5 is discussed in the last section.

3 The Suita conjecture and L^2-extension

Let D be a (say smoothly) bounded domain in $\mathbb{C}$ and suppose $0 \in D$. The Bergman space of D is

$$A^2(D) = \{h \in H(D); \|h\|^2 := \int_D |h|^2 d\lambda < \infty\},$$

and the Bergman kernel at 0 is

$$B(0) = \sup_{\|h\| \leq 1} |h(0)|^2.$$

To state Suita's conjecture we also need the Green's function $G(z)$ which is a subharmonic function in D, vanishing at the boundary, with a logarithmic pole at 0;

$$G(z) = \log|z|^2 - v(z),$$

where v is harmonic and chosen so that G vanishes at the boundary. By definition, $v(0) := c_D(0) = c_D$ is the Robin constant of D at 0. Suita's conjecture, which was proved in

Błocki [2013], Guan and Zhou [2015], says that

$$B(0) \geq \frac{e^{-c_D}}{\pi}.$$

When D is a disk with center 0 it is clear that equality holds, and it was proved in Guan and Zhou [ibid.] that equality holds only then. In Błocki [2015], Błocki gave a much simpler proof of the conjecture, based on variations of domains and the 'tensor power trick'. In connection with this, László Lempert proposed an even simpler proof, using (pluri)subharmonic variation of Bergman kernels from Maitani and Yamaguchi [2004], Berndtsson [2006], that we shall now describe (see Błocki [2015], Berndtsson and Lempert [2016]).

Let for $t \leq 0$,

$$D_t = \{z \in D; G(z) < t\},$$

and $B_t(0)$ be the Bergman kernel for D_t. Since

$$\mathfrak{D} := \{(\tau, z) \in \mathbb{C} \times D; G(z) - \operatorname{Re} \tau < 0\}$$

is a pseudoconvex domain in $\mathbb{C}^2$, and $D_t = \mathfrak{D}_\tau$, the vertical slice of $\mathfrak{D}$, if $t = \operatorname{Re} \tau$, it follows from Maitani and Yamaguchi [2004], Berndtsson [2006] (cf Theorem 2.4) that $\log B_t(0)$ is convex. When t approaches $-\infty$, D_t is very close to a disk centered at the origin with radius $e^{(t+c_D)/2}$, so

$$B_t \sim \frac{e^{-t-c_D}}{\pi}.$$

In particular, the function

$$k(t) := \log B_t(0) + t$$

is convex and bounded on the negative half axis. Therefore, it must be increasing. Hence

$$B(0) = e^{k(0)} \geq \lim_{t \to -\infty} e^{k(t)} = \frac{e^{-c_D}}{\pi},$$

which proves Suita's conjecture.

It is clear that similar proofs work in higher dimensions and also for weighted Bergman spaces with a plurisubharmonic weight function. The main new observation in Berndtsson and Lempert [2016] is that the same technique can be used to give a proof of the Ohsawa-Takegoshi extension theorem, Ohsawa and Takegoshi [1987].

We first recall a simple version of this important result. We let D be a bounded pseudoconvex domain in $\mathbb{C}^n$ and ϕ a plurisubharmonic function in D. Let V be a linear subspace of $\mathbb{C}^n$ of codimension m which intersects D. The Ohsawa-Takegoshi theorem says that in this situation (and actually under much more general conditions), for any holomorphic function f on $V \cap D$, there is a holomorphic function F in D, such that $f = F$ on V and

$$\int_D |F|^2 e^{-\phi} d\lambda_n \leq C \int_{V\cap D} |f|^2 e^{-\phi} d\lambda_{n-m},$$

where C is a constant depending only on the diameter of D. What makes this theorem so powerful is that even though there is no assumption of strict plurisubharmonicity or smoothness, we get an estimate with an absolute constant. Note that, when V is just a point, we get the existence of a function F in D with good L^2-estimates and prescribed value at the origin; this is equivalent to an estimate for the Bergman kernel.

Just as in Suita's conjecture, one can now ask for optimal estimates. Such estimates were given in Błocki [2013] and more generally by Guan and Zhou [2015]. We shall now see that results of this type also follow from Theorem 2.3. We also point out that, conversely, Guan and Zhou show that results along the line of Theorem 2.3 and Theorem 2.5 follow from sharp versions of the Ohsawa-Takegoshi extension theorem, so in very general terms the two types of results are perhaps 'equivalent'.

Write for z in $\mathbb{C}^n$, $z = (z_1, ...z_m, z_{m+1}, ...z_n) = (z', z'')$, and say that V is defined by the equation $z' = 0$. Let

$$G(z) = \log |z'|^2,$$

and let

$$D_t = \{z \in D; G(z) < t\}.$$

We may assume that $|z'| \leq 1$ in D. For each $t \in (-\infty, 0)$ we let F_t be the extension of f to D_t of minimal norm. Such a minimal element exists by general Hilberts space theory. The following result is not explicitly stated in Berndtsson and Lempert [2016], but implicitly contained there.

Proposition 3.1.

$$e^{-mt}\|F_t\|_t^2 = e^{-mt} \int_{D_t} |F_t|^2 e^{-\phi} d\lambda$$

is a decreasing function of t.

It follows that

$$\|F_0\|^2 \leq \lim_{t\to-\infty} e^{-mt}\|F_t\|_t^2.$$

It is not hard to see that the last limit equals (with σ_m the volume of the m-dimensional unit ball)

$$\sigma_m \int_{V\cap D} |f|^2 e^{\phi} d\lambda_{n-m},$$

at least when ϕ is smooth in a neighbourhood of the closure of D. This gives the sharp extension theorem in this setting, and the general case is obtained by the usual approximation procedures.

Proposition 3.1 can be proved much like the Suita conjecture, but using the general Theorem 2.3 instead of plurisubharmonic variation of Bergman kernels. First note that since any function f on $V \cap D$ extends to D, the space of holomorphic functions on $V \cap D$ can be identified with the quotient space $H(D)/J(V)$, or $H(D_t)/J(V)$, where $J(V)$ is the subspace of functions that vanish on V. The quantity that we want to estimate, $\|F_t\|_t$, is the norm of f in $A^2(D_t)/J(V)$. The space of measures with compact support in $V \cap D$ is dense in the dual of $H(D_t)/J(V)$. We first prove a dual estimate on such measures:

$$\|\mu\|_t^2 e^{mt}$$

is increasing, with

$$\|\mu\|_t^2 = \sup_F \frac{|\int F d\mu|^2}{\int_{D_t} |F|^2 e^{-\phi} d\lambda}.$$

This is proved in almost exactly the same way as in the Suita case, when V is a point, and Proposition 3.1 follows.

4 The space of Kähler metrics

In this section we will look at fibrations with $X = Z \times U$ where Z is a compact complex manifold, U is a domain in $\mathbb{C}$ and $p : X \to U$ is the projecion on the second factor. We also assume given a complex line bundle, L over Z. Pulling it back to X by the projection on the first factor we get a line bundle over X, that we denote by the same letter, sometimes writing $L \to X$ or $L \to Z$, to make the meaning clear. Mostly we assume that $L \to Z$ is positive, in the sense that it has some smooth metric of positive curvature, but we don't fix any metric – instead we will use this set up to study the space M_L of positively curved metrics on L.

Here all the fibers $X_t = Z$ are the same and the line bundles $L|_{X_t}$ are also the same. A metric on $L \to X$ can be interpreted as a complex curve of metrics on L, parametrized by τ in U. Of particular interest for us is the case when $U = \{\tau \in \mathbb{C}; 0 < \operatorname{Re}\tau < 1\}$ is a strip and ϕ only depends on the real part of τ; then we can interpret ϕ as a real curve in M_L.

The space M_L was introduced by Mabuchi [1987]. It is an open subset of an affine space modeled on $C^\infty(Z)$, therefore the tangent space of M_L at any point ϕ is naturally identified with $C^\infty(Z)$. Mabuchi defined a Riemannian structure on M_L by

$$\|\chi\|_\phi^2 := \int_Z |\chi|^2 \omega_\phi^n,$$

for $\chi \in C^\infty(Z)$, with $\omega_\phi := i\partial\bar\partial\phi$, see also Semmes [1988] and S. K. Donaldson [1999]. As proved in these papers, a curve in M_L, i.e. a metric on L depending only on $t = \operatorname{Re}\tau$

is a geodesic for the Riemannian metric if and only if $i\partial\bar\partial\phi \geq 0$ on X, and it satisfies the homogenous complex Monge-Ampère equation

$$(i\partial\bar\partial\phi)^{n+1} = 0.$$

In our terminology in Section 2, this means that $c(\phi) = 0$, and in general, $c(\phi)$ can be interpreted as the geodesic curvature of the curve. Since honest geodesics are scarce, see Lempert and Vivas [2013], Darvas and Lempert [2012], Ross and Nyström [2015], Ross and Nyström [n.d.] for problems encountered in solving the homogenous complex Monge-Ampère equation involved, we will also use 'generalized geodesics'. These are metrics on L as above that are only assumed to be locally bounded, such that $i\partial\bar\partial\phi \geq 0$ and $(i\partial\bar\partial\phi)^{n+1} = 0$ in the sense of pluipotential theory (Bedford and Taylor [1976]). It is easy to see (Berndtsson [2011]) that any two metrics in M_L can be connected with a generalized geodesic in this sense, and by a famous result of Chen [2000b], Janeczko [1996], the geodesic has in fact higher regularity, so that $\partial\bar\partial\phi$ is a current with bounded coefficients. We will say that such geodesics are of class $C^{(1,1)}$. There is by now an extensive theory on these matters, for which we refer to Mabuchi [1987], Semmes [1988], S. K. Donaldson [1999], Phong and Sturm [2009] and S. K. Donaldson [2005].

We first consider the case when $L = -K_Z$ is the anticanonical bundle of Z. The positivity of L then means that Z is a Fano manifold. The fibers of the vector bundle E, defined in Section 2,

$$E_\tau = H^0(X_\tau, K_{X_\tau} + L) = H^0(Z, \mathbb{C})$$

are then just the space of constant functions on Z (so E is a line bundle). A metric $e^{-\phi}$ on $L \to Z$ can be identified with a volume form on Z that we also denote $e^{-\phi}$, and with this somewhat abusive notation we have for the element '1' in $H^0(Z, \mathbb{C})$,

$$\|1\|^2_\phi = \int_Z e^{-\phi}.$$

We are therefore in the situation described in the introduction when E is a line bundle with a nonvanishing section. Therefore we get a Prékopa type theorem for Fano manifolds.

Theorem 4.1. *(Berndtsson [2015]) Let $e^{-\phi}$ be a locally bounded metric on $L \to Z \times U$, where $L \to Z$ is the anticanonical bundle of Z. Assume that U is a strip and that ϕ depends only on the real part of $\tau \in U$. Suppose also that $i\partial\bar\partial\phi \geq 0$ in the sense of currents. Let*

$$\tilde\phi(t) = -\log \int_Z e^{-\phi(t,\cdot)}.$$

Then $\tilde\phi$ is a convex function of t. If $\tilde\phi$ is linear, then there is a holomorphic vector field V on U, which is a lift of $\partial/\partial\tau$ on U, whose flow maps fibers X_τ to fibers and $i\partial\bar\partial\phi|_{X_\tau}$ to $i\partial\bar\partial\phi|_{X_\tau}$.

When ϕ is smooth the first part of the theorem follows immediately from Theorem 2.5, and the general case is easily obtained by approximation. The second part is clearly of the same vein as Corollary 2.7, but it does not follow directly from there because of the lack of smoothness and strict positivity.

The main interest of this result lies in its connection with Kähler-Einstein metrics. To explain this we also need to introduce the Monge-Ampère energy, $\mathcal{E}(\phi)$, of a metric. This is a real valued function on M_L, which can be defined (up to a constant) by the property that for any curve $t \to \phi_t \in M_L$,

$$\frac{d}{dt}\mathcal{E}(\phi_t) = \frac{1}{Vol(L)} \int_Z \dot{\phi}_t \omega_{\phi_t}^n / n!,$$

where $\dot{\phi} = d\phi/dt$ and the volume of the line bundle is

$$Vol(L) := \int_Z \omega_\phi^n / n!$$

(it does not depend on the metric). One then defines the Ding functional, Ding [1988], as

$$D(\phi) := \log \int_Z e^{-\phi} + \mathcal{E}(\phi).$$

The critical points of the Ding functional are the metrics that satisfy

$$e^{-\phi} = C\omega_\phi^n.$$

This means precisely that the Ricci curvature of the Kähler metric ω_ϕ equals ω_ϕ, i.e. that ω_ϕ is a Kähler-Einstein metric (of positive curvature). It is well known that the Monge-Ampère energy is linear along (even generalized) geodesics, so the Ding functional is linear along a geodesic precisely when its first term

$$t \to \log \int_Z e^{-\phi_t}$$

is linear. As noted by Berman [n.d.], this gives a proof of the Bando-Mabuchi uniqueness theorem (Bando and Mabuchi [1987]), for Kähler-Einstein metrics on Fano manifolds. Indeed, suppose $e^{-\phi_0}$ and $e^{-\phi_1}$ are metrics on the anticanonical bundle of Z, and that $i\partial\bar{\partial}\phi_0$ and $i\partial\bar{\partial}\phi_1$ are both Kähler-Einstein. Connect them by a generalized geodesic. Since both endpoints are critical points for the Ding functional, and the Ding functional is concave along the geodesic, it must in fact be linear. Then Theorem 4.1 implies that ω_{ϕ_0} and ω_{ϕ_1} are connected via the flow of a holomorphic vector field, i.e. an element in the identity component of the automorphism group. This is the Bando-Mabuchi theorem.

As also noted by Berman [n.d.], similar arguments prove a variant of the Moser-Trudinger-Onofri inequality (Moser [1970/71], Onofri [1982]) for Kähler-Einstein Fano manifolds (the original case of the theorem was for Z equal to the Riemann sphere). The (variant of) the theorem says, in our terminology, that Kähler-Einstein metrics are global maxima for the Ding functional on the space of all positively curved metrics, which is clear from the concavity, since Kähler-Einstein metrics are critical points. The original version of the theorem, for the Riemann sphere, deals with metrics that are not necessarily positively curved, but as shown by Berman, this follows in one dimension from the positively curved case by an elegant trick.

The formalism described here can easily be generalized. We fix a metric $e^{-\psi}$ on a pseudoeffective $\mathbb{R}$-line bundle L' and study metrics $e^{-\phi}$ on an $\mathbb{R}$-line bundle L'', such that $L' + L'' = L = -K_Z$. With this we define the twisted Ding functional

$$D_\psi(\phi) := \log \int_Z e^{-\phi-\psi} + \mathcal{E}(\phi),$$

where $\mathrm{Vol}(L)$ is replaced by $\mathrm{Vol}(L'')$ in the definition of $\mathcal{E}$. The critical points of D_ψ satisfy

$$e^{-\phi-\psi} = C\omega_\phi^n.$$

Hence the Ricci curvature of ω_ϕ satisfies

$$Ric(\omega_\phi) = \omega_\phi + i\partial\bar\partial\psi,$$

so ω_ϕ is a 'twisted Kähler-Einstein metric'. It is clear that if ψ is assumed to be bounded, the same argument as before gives uniqueness modulo automorphisms, but actually the argument can be generalized to when $i\partial\bar\partial\psi = \beta[\Delta]$ is a multiple of a current defined by a divisor in Z. This leads to a uniqueness theorem for Kähler-Einstein metrics with conical singularities, introduced by S. K. Donaldson [2012], as a tool for the solution of the Yau-Tian-Donaldson conjecture, see Chen, S. Donaldson, and S. Sun [2014] and subsequent papers. In these papers it was also shown that this generalized version of Theorem 4.1 can be used to prove reductivity of the group of automorphisms of the pair (Z, Δ). Furthermore, the theorem was generalized to 'log-Fano' manifolds, that arise naturally in this context, in Eyssidieux, Guedj, and Zeriahi [2009]. In this context, see also Berman [2016] for applications to the converse direction of the Yau-Tian-Donaldson conjecture.

So far we have discussed only the case when L is the anticanonical bundle of Z, and the resulting convexity preperties of the Ding functional, but it turns out that the formalism is also useful in connection with other functionals in Kähler geometry. A case in point is the Mabuchi K-energy, Mabuchi [1986], $\mathfrak{M}$. The K-energy of a metric $e^{-\phi}$ on a positive line bundle L can be defined (again up to a constant) by

$$\frac{d}{dt}\mathfrak{M}(\phi_t) = \int_Z \dot{\phi}(S_\phi - \hat{S}_\phi)\omega_\phi^n/n!,$$

where $t \to \phi_t$ is any smooth curve in M_L, S_ϕ is the scalar curvature of the Kähler metric ω_ϕ, and $\hat{S}_\phi$ is the average of S_ϕ over Z. The raison d'être of $\mathfrak{M}$ is that its critical points are precisely (the potentials of) the metrics of constant scalar curvature. It was proved by Mabuchi [1986], that $\mathfrak{M}$ is convex along smooth geodesics, but, again, in applications there is a need to consider also generalized geodesics. It was proved by Chen [2000a], that the K-energy can be defined along any generalized geodesic of class $C^{(1,1)}$, which is crucial since any two points in the space can be connected by such geodesics. (For this, he rewrites the definition of $\mathfrak{M}$ since the original definition involves four derivatives of ϕ.) Chen also conjectured that $\mathfrak{M}$ would be convex along any generalized geodesic of class $C^{(1,1)}$. This was proved in a joint paper with Berman and Berndtsson [2017].

Theorem 4.2. *The Mabuchi K-energy is convex along any generalized geodesic of class $C^{(1,1)}$.*

(An alternative proof was later given in Chen, Li, and Păuni [2016].) Using this we proved the uniqueness of metrics of constant scalar curvature up to flows of holomorphic vector fields:

Theorem 4.3. *Let Z be a compact manifold and $L \to Z$ a positive line bundle. Let ω_0 and ω_1 be two Kähler metrics on of constant scalar curvature in $c_1[L]$. Then there is a holomorphic vector field on Z, with time 1 flow F, such that $F^*(\omega_1) = \omega_0$.*

Uniqueness was proved earlier, in case Z has discrete automorphism group in S. K. Donaldson [2001]. Our result is in fact more general; it treats not only metrics of constant scalar curvature, but general 'extremal metrics', and does not assume that the cohomology class of the metrics is integral. For this, and a discussion of previous work, we refer to Berman and Berndtsson [2017].

5 Variation of complex structure

In this section we will mainly consider families of canonically polarized manifolds. We assume that $p : X \to U$ is a smooth proper fibration, that U is a domain in $\mathbb{C}$, and that the fibers X_t have positive canonical bundle. In this setting, X is automatically Kähler, since we may find a smooth metric $e^{-\psi}$ on $K_{X/U}$ which is positively curved on fibers, and take a Kähler form as $i\partial\bar{\partial}\psi + p^*(\omega)$ where ω is sufficiently positive on the base. Hence the results from Section 2 apply.

By the Aubin-Yau theorem, S. T. Yau [1978], Aubin [1982], each fiber has a unique Kähler-Einstein metric (now with Ricci curvature equal to -1). This metric can be written

$$\omega = i\partial\bar\partial\phi,$$

where $e^{-\phi}$ is a metric on the canonical bundle. Then e^{ϕ} is a metric on the anticanonical bundle, which as in the previous section can be identified with a volume form, and the metric is unique if we normalize so that

$$e^{\phi} = (\omega_\phi)^n/n!.$$

Applying this to each fiber X_t we get a metric on the relative canonical bundle $K_{X/U}$ that we also denote by $e^{-\phi}$. An important theorem of Schumacher [2012] implies that $i\partial\bar\partial\phi$ is semipositively curved not only along the fibers, but on the total space:

Theorem 5.1. *If $e^{-\phi}$ is the normalized Kähler-Einstein potential described above, and $c(\phi)$ is defined as in Section 2, then*

$$\Box c(\phi) + c(\phi) = |\kappa_\phi|^2$$

on each fiber. As a consequence, $c(\phi)$ is semipositive and strictly positive on each fiber where κ_ϕ does not vanish identically.

Here $\Box = \bar\partial^*\bar\partial$ is the Laplace operator on functions on a fiber, for the Kähler metric $\omega_\phi|_{X_t}$, and κ_ϕ, as defined as in Section 2 turns out to be the harmonic representative of the Kodaira-Spencer class. The non negativity part of the statement follows immediately from the differential equation, since $\Box c(\phi) \leq 0$ at a minimum point, and the strict positivity part is also a well known property of elliptic equations. Since $i\partial\bar\partial\phi > 0$ along the fibers, the positivity of $c(\phi)$ implies that $i\partial\bar\partial\phi$ is positive on the total space.

This result can be seen as an analog of Theorem 2.8, when L is trivial or a power of the relative canonical bundle. Indeed, that theorem says the Bergman kernel defines a semipositive metric on the relative canonical bundle, at least if it is not identically zero. Tsuji [2010], independently proved the semipositivity part of Theorem 5.1, using Theorem 2.8 and iteration:

Theorem 5.2. *Let X be a family of canonically polarized manifolds. For any sufficiently positive line bundle $(L, e^{-\psi})$ on X, let $e^{-s(\psi)}$ be the Bergman kernel metric on $K_{X/U}+L$. Define iteratively a sequence of metrics $h_m = e^{-s^m(\psi)}$ on $mK_{X/U} + L$ in this way. Then an appropriate renormalization of $h_m^{1/m}$ tends to the metric on $K_{X/U}$ defined by the Kähler-Einstein potentials.*

Since all the metrics in the iteration have semipositive curvature by Theorem 2.8, this gives a different proof of the semipositivity part of Theorem 5.1. Tsuji also applied these arguments to situations when we assume much less positivity along the fibers.

Both Schumacher's theorem and Theorem 2.8 give some positivity of $K_{X/Y}$ under assumptions of fiberwise positivity. Schumacher's theorem was generalized by Paun [n.d.], to twisted relative canonical bundles, and even to general adjoint classes, not necessarily integral. He also applied this result to solve a long standing conjecture about the surjectivity of the Albanese map for compact Kähler manifolds with nef anticanonical bundle.

Theorem 5.3. *Let $p : X \to Y$ be a holomorphic surjective map between compact Kähler manifolds. Let β be a semipositive closed $(1,1)$-form on on X and assume that the cohomology class $c_1[K_{X/Y}] + [\beta]$ contains a $(1,1)$-form Ω which is strictly positive on all fibers X_y, for y outside a proper analytic subset, Y_0, of Y. Then $c_1[K_{X/Y}] + \beta$ contains a closed semipositive current, which is smooth outside $p^{-1}(Y_0)$.*

Let us now consider the particular case of a fibration $p : X \to U$ where the fibers are Riemann surfaces of a certain genus $g \geq 2$. If we choose $L = K_{X/U}$, our vector bundle E with fibers

$$E_t = H^{1,0}(X_t, K_{X/Y}) = H^0(X_t, 2K_{X/U})$$

is the dual of the bundle with fibers

$$E_t^* = H^{0,1}(X_t, T^{1,0}(X_t)),$$

which is the space of infinitesimal deformations of the complex structure on X_t. Any positively curved metric on E therefore induces a negatively curved metric on the space of deformations (along U). In particular, taking the L^2-metric on E, defined by the metric on $K_{X/U}$ given by the Kähler-Einstein potentials, $e^{-\phi}$, we get the Weil-Peterson metric on U. By Schumacher's theorem, $e^{-\phi}$ is (semi)positively curved on X, so we get the classical fact that the Weil-Peterson metric has seminegative curvature, and negative curvature where the Kodaira-Spencer class does not vanish, Ahlfors [1961/1962], Royden [1975], Wolpert [1986]. Moreover, from Theorem 2.6, we get an explicit formula for the curvature. Using Schumacher's theorem again (now the differential equation for $c(\phi)$), we can rewrite the formula so that it coincides with the formula found by Wolpert [ibid.], see Berndtsson [2011]. The same argument works when U has higher dimension, and then Theorem 2.5 shows that the Weil-Peterson metric has dual Nakano-negative curvature, cf Liu, X. Sun, and S.-T. Yau [2009]. This is stronger than negative bisectional curvature, which corresponds to Griffiths negativity.

The case of families of higher dimensional canonically polarized manifolds is significantly more complicated. We are then primarily interested in vector bundles $\mathcal{H}^{1,n-1}$ with fibers

$$H^{1,n-1}(X_t, K_{X/U})$$

and their positivity properties. The reason for this is that the dual of $H^{1,n-1}(X_t, K_{X/U}) = H^{n,n-1}(X_t, (T^*)^{1,0}(X_t))$ is $H^{0,1}(X_t, T^{1,0})$, the space of infinitesimal deformations of the complex structure on X_t. The Kodaira-Spencer map sends the tangent bundle of U into this space, and therefore any metric on $\mathcal{H}^{1,n-1}$ induces a metric on the base. The Weil-Peterson metric on the base (which can be obtained this way) was first studied by Siu [1986], who found an explicit formula for its curvature. This was generalized by Schumacher [2012], who found a curvature formula for the bundle $\mathcal{H}^{1,n-1}$ and also for the other higher direct image bundles $\mathcal{H}^{p,q}$, with $p+q = n$. Both Siu's and Schumacher's formulas contain terms that give an apparent positive contribution to the curvature of the Weil-Peterson metric, but as shown in Schumacher [ibid.] and To and Yeung [2015], the metrics on all the higher direct images bundles can be combined to give a Finsler metric of strictly negative sectional curvature that can partly substitute for the Weil-Peterson metric.

All these works focus on the relative canonical bundle and the metric on it given by the Kähler-Einstein potential, and in the case of one dimensional fibers they are special cases Theorem 2.6. The generalization to general line bundles L with metrics of fiberwise positive curvature was obtained in Naumann [n.d.] and Berndtsson, Paun, and Wang [n.d.]. In Berndtsson, Paun, and Wang [ibid.], the formalism was also extended to families of Calabi-Yau manifolds (where related results have also been announced by To and Yeung). Finally, we also mention that, building on work by Lu [2001], Wang [n.d.], has found a different approach to these problems, which so far works for Calabi-Yau families and then produces a Hermitian version of the Weil-Peterson metric, with negative bisectional curvature. The main idea is to embed the tangent bundle of U, not in the dual of $\mathcal{H}^{1,n-1}$ as above, but in the endomorphism bundle of the sum of all the $\mathcal{H}^{p,q}$.

6 Positivity of direct image sheaves

We shall finally discuss extensions of Theorem 2.5 under less restrictive assumptions. Let $p : X \to Y$ be a surjective holomorphic map between two projective varieties, and let $L \to X$ be a holomorphic line bundle equipped with a singular metric $e^{-\phi}$ with semi-positive curvature current $i\partial\bar{\partial}\phi \geq 0$. The 'multiplier ideal sheaf' $\mathcal{I}(\phi)$ is the sheaf of holomorphic functions on X that are square integrable against $e^{-\phi}$. The first problem is that under these general circumstances we no longer get a vector bundle on the base. The role of the vector bundle E is instead played by the (zeroth) direct image sheaf,

$$\mathcal{E} := p_*((K_{X/Y} + L) \otimes \mathcal{I}(\phi)).$$

$\mathcal{E}$ is a sheaf over Y whose sections over an open set U in the base are the sections of $(K_{X/Y} + L) \otimes \mathcal{I}(\phi)$ over $p^{-1}(U)$. By a classical theorem of Grauert, $\mathcal{E}$ is coherent and it is also torsion free. The consequence of this that we will use is that it is locally free outside of a subvariety, Y_0, of Y of codimension at least two. This means that outside of Y_0, $\mathcal{E}$ is the sheaf of sections of a vector bundle E. In the setting of Theorem 2.5, Y_0 is empty and $\mathcal{E}$ coincides with the sheaf of sections of E as defined there. In general, let Y_1 be a proper subvariety of Y such that p is a submersion outside $p^{-1}(Y_1)$. We can now define a L^2-metric on E over $p^{-1}(Y \setminus (Y_0 \cup Y_1))$ as we did when $e^{-\phi}$ was assumed to be smooth. The new feature that appears is that this is now a singular metric. For singular metrics on a vector bundle it seems impossible to define positivity and negativity in terms of a curvature current (cf Raufi [2015]), but we can circumvent this problem in a way similar to what we did in Theorem 2.3: We say that a singular metric is negatively curved if the logarithm of the norm of any holomorphic section is plurisubharmonic, and it is positively curved if the dual is negative (cf Berndtsson and Păun [2008]).

Paun and Takayama extended the notion of singular metrics on vector bundles to the setting of coherent, torsion free sheaves. The definition is simply that a singular metric on such a sheaf is a singular metric on the vector bundle defined by the sheaf outside of Y_0. Since negativity (and positivity) of the curvature is defined in terms of plurisubharmonic functions, and since plurisubharmonic functions extend over varieties of codimension at least 2, it turns out that this is a useful definition. Given all this, we have the following theorem of Păun and Takayama [n.d.], and Hacon, Popa, and Schnell [n.d.], which seems to be the most general theorem on (metric) positivity of direct images.

Theorem 6.1. *The L^2-metric on $E = p_*((K_{X/Y} + L) \otimes \mathcal{I}(\phi))$ over the complement of $Y_0 \cup Y_1$ extends to a (singular) metric of nonnegative curvature on the coherent and torsion free sheaf $\mathcal{E}$.*

(Paun and Takayama proved this when $\mathcal{I}(\phi)$ is trivial on a generic fiber.) The next theorem is also due to Paun and Takayama:

Theorem 6.2. *Any coherent torsion free sheaf which has a (singular) metric of nonnegative curvature is weakly positive in the sense of Viehweg [1995]. Hence, by Theorem 6.1,*

$$\mathcal{E} := p_*((K_{X/Y} + L) \otimes \mathcal{I}(\phi))$$

is weakly positive.

In very general terms, the notion of 'positivity' in algebraic geometry, is often given in terms of the existence of certain holomorphic or algebraic objects. Thus, e. g. the positivity of a line bundle means that a high power of it has enough sections to give an embedding into projective space. Similarily, Viehweg's weak positivity of a sheaf $\mathcal{F}$ means that certain associated sheaves are generated by global sections. From the analytic point of view, the existence of such sections should be a consequence of the metric positivity of curvature, and the previous theorem gives a very general version of this. We will not go into more details on these matters or give the exact definition of positivity in the sense of Viehweg – it would require another article (and another author.)

Instead we end by a major application of these results, due to Cao and Păun [2017].

Theorem 6.3. *Let $p : X \to A$ be a surjective holomorphic map, where X is smooth projective and A is an Abelian variety. Then we have the inequality for the Kodaira dimensions of X and a generic fiber, F,*

$$\kappa(X) \geq \kappa(F).$$

This proves the case of the famous 'Itaka conjecture', $\kappa(X) \geq \kappa(F) + \kappa(Y)$, when Y is an Abelian variety (and hence has Kodaira dimension zero). For a simplification of the proof and a generalization of the result we refer to Hacon, Popa, and Schnell [n.d.], which is also a beautiful survey of the field.

References

Lars V. Ahlfors (1961/1962). "Curvature properties of Teichmüller's space". *J. Analyse Math.* 9, pp. 161–176. MR: 0136730 (cit. on p. 896).

Thierry Aubin (1982). *Nonlinear analysis on manifolds. Monge-Ampère equations*. Vol. 252. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, New York, pp. xii+204. MR: 681859 (cit. on p. 895).

Shigetoshi Bando and Toshiki Mabuchi (1987). "Uniqueness of Einstein Kähler metrics modulo connected group actions". In: *Algebraic geometry, Sendai, 1985*. Vol. 10. Adv. Stud. Pure Math. North-Holland, Amsterdam, pp. 11–40. MR: 946233 (cit. on p. 892).

Eric Bedford and B. A. Taylor (1976). "The Dirichlet problem for a complex Monge-Ampère equation". *Invent. Math.* 37.1, pp. 1–44. MR: 0445006 (cit. on p. 891).

Robert J. Berman (n.d.). "Sharp inequalities for determinants of Toeplitz operators and dbar-Laplacians on line bundles". arXiv: 0905.4263 (cit. on pp. 892, 893).

– (2016). "K-polystability of $\mathbb{Q}$-Fano varieties admitting Kähler-Einstein metrics". *Invent. Math.* 203.3, pp. 973–1025. MR: 3461370 (cit. on p. 893).

Robert J. Berman and Bo Berndtsson (2017). "Convexity of the K-energy on the space of Kähler metrics and uniqueness of extremal metrics". *J. Amer. Math. Soc.* 30.4, pp. 1165–1196. MR: 3671939 (cit. on p. 894).

Bo Berndtsson (2006). "Subharmonicity properties of the Bergman kernel and some other functions associated to pseudoconvex domains". *Ann. Inst. Fourier (Grenoble)* 56.6, pp. 1633–1662. MR: 2282671 (cit. on pp. 883, 884, 888).

– (2009). "Curvature of vector bundles associated to holomorphic fibrations". *Ann. of Math. (2)* 169.2, pp. 531–560. MR: 2480611 (cit. on pp. 881, 882, 885).

Bo Berndtsson (2011). "Strict and nonstrict positivity of direct image bundles". *Math. Z.* 269.3-4, pp. 1201–1218. MR: 2860284 (cit. on pp. 885, 886, 891, 896).

– (2015). "A Brunn-Minkowski type inequality for Fano manifolds and some uniqueness theorems in Kähler geometry". *Invent. Math.* 200.1, pp. 149–200. MR: 3323577 (cit. on p. 891).

Bo Berndtsson and László Lempert (2016). "A proof of the Ohsawa-Takegoshi theorem with sharp estimates". *J. Math. Soc. Japan* 68.4, pp. 1461–1472. MR: 3564439 (cit. on pp. 888, 889).

Bo Berndtsson and Mihai Păun (2008). "Bergman kernels and the pseudoeffectivity of relative canonical bundles". *Duke Math. J.* 145.2, pp. 341–378. MR: 2449950 (cit. on pp. 887, 898).

Bo Berndtsson, Mihai Paun, and Xu Wang (n.d.). "Algebraic fiber spaces and curvature of higher direct images". arXiv: 1704.02279 (cit. on p. 897).

Zbigniew Błocki (2013). "Suita conjecture and the Ohsawa-Takegoshi extension theorem". *Invent. Math.* 193.1, pp. 149–158. MR: 3069114 (cit. on pp. 888, 889).

– (2015). "Bergman kernel and pluripotential theory". In: *Analysis, complex geometry, and mathematical physics: in honor of Duong H. Phong*. Vol. 644. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 1–10. MR: 3372456 (cit. on p. 888).

Herm Jan Brascamp and Elliott H. Lieb (1976). "On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation". *J. Functional Analysis* 22.4, pp. 366–389. MR: 0450480 (cit. on p. 878).

Junyan Cao and Mihai Păun (2017). "Kodaira dimension of algebraic fiber spaces over abelian varieties". *Invent. Math.* 207.1, pp. 345–387. MR: 3592759 (cit. on p. 899).

Xiuxiong Chen (2000a). "On the lower bound of the Mabuchi energy and its application". *Internat. Math. Res. Notices* 12, pp. 607–623. MR: 1772078 (cit. on p. 894).

– (2000b). "The space of Kähler metrics". *J. Differential Geom.* 56.2, pp. 189–234. MR: 1863016 (cit. on p. 891).

Xiuxiong Chen, Simon Donaldson, and Song Sun (2014). "Kähler-Einstein metrics and stability". *Int. Math. Res. Not. IMRN* 8, pp. 2119–2125. MR: 3194014 (cit. on p. 893).

XiuXiong Chen, Long Li, and Mihai Păuni (2016). "Approximation of weak geodesics and subharmonicity of Mabuchi energy". *Ann. Fac. Sci. Toulouse Math. (6)* 25.5, pp. 935–957. MR: 3582114 (cit. on p. 894).

Dario Cordero-Erausquin (2005). "On Berndtsson's generalization of Prékopa's theorem". *Math. Z.* 249.2, pp. 401–410. MR: 2115450 (cit. on p. 878).

Dario Cordero-Erausquin and Bo'az Klartag (2012). "Interpolations, convexity and geometric inequalities". In: *Geometric aspects of functional analysis*. Vol. 2050. Lecture Notes in Math. Springer, Heidelberg, pp. 151–168. MR: 2985131 (cit. on p. 878).

Tamás Darvas and László Lempert (2012). "Weak geodesics in the space of Kähler metrics". *Math. Res. Lett.* 19.5, pp. 1127–1135. MR: 3039835 (cit. on p. 891).

Jean-Pierre Demailly (1997). *Complex analytic and differential geometry. Notes available on Demailly's webpage*. Citeseer (cit. on pp. 878, 882).

Wei Yue Ding (1988). "Remarks on the existence problem of positive Kähler-Einstein metrics". *Math. Ann.* 282.3, pp. 463–471. MR: 967024 (cit. on p. 892).

S. K. Donaldson (1999). "Symmetric spaces, Kähler geometry and Hamiltonian dynamics". In: *Northern California Symplectic Geometry Seminar*. Vol. 196. Amer. Math. Soc. Transl. Ser. 2. Amer. Math. Soc., Providence, RI, pp. 13–33. MR: 1736211 (cit. on pp. 890, 891).

– (2001). "Scalar curvature and projective embeddings. I". *J. Differential Geom.* 59.3, pp. 479–522. MR: 1916953 (cit. on p. 894).

– (2005). "Scalar curvature and projective embeddings. II". *Q. J. Math.* 56.3, pp. 345–356. MR: 2161248 (cit. on p. 891).

– (2012). "Kähler metrics with cone singularities along a divisor". In: *Essays in mathematics and its applications*. Springer, Heidelberg, pp. 49–79. MR: 2975584 (cit. on p. 893).

Serge Dubuc (1977). "Critères de convexité et inégalités intégrales". *Ann. Inst. Fourier (Grenoble)* 27.1, pp. x, 135–165. MR: 0444863 (cit. on p. 886).

Philippe Eyssidieux, Vincent Guedj, and Ahmed Zeriahi (2009). "Singular Kähler-Einstein metrics". *J. Amer. Math. Soc.* 22.3, pp. 607–639. MR: 2505296 (cit. on p. 893).

Takao Fujita (1978). "On Kähler fiber spaces over curves". *J. Math. Soc. Japan* 30.4, pp. 779–794. MR: 513085 (cit. on p. 880).

R. J. Gardner (2002). "The Brunn-Minkowski inequality". *Bull. Amer. Math. Soc. (N.S.)* 39.3, pp. 355–405. MR: 1898210 (cit. on p. 877).

Phillip Griffiths and Loring Tu (1984). "Curvature properties of the Hodge bundles". In: *Topics in transcendental algebraic geometry (Princeton, N.J., 1981/1982)*. Vol. 106. Ann. of Math. Stud. Princeton Univ. Press, Princeton, NJ, pp. 29–49. MR: 756844 (cit. on p. 880).

Qi'an Guan and Xiangyu Zhou (2015). "A solution of an L^2 extension problem with an optimal estimate and applications". *Ann. of Math. (2)* 181.3, pp. 1139–1208. MR: 3296822 (cit. on pp. 888, 889).

Christopher Hacon, Mihnea Popa, and Christian Schnell (n.d.). "Algebraic fiber spaces over abelian varieties: around a recent theorem by Cao and Paun". arXiv: 1611.08768 (cit. on pp. 898, 899).

Lars Hörmander (1965). "L^2 estimates and existence theorems for the $\bar\partial$ operator". *Acta Math.* 113, pp. 89–152. MR: 0179443 (cit. on p. 878).

S. Janeczko (1996). "Bifurcations of the center of symmetry". *Geom. Dedicata* 60.1, pp. 9–16. MR: 1376477 (cit. on p. 891).

Yujiro Kawamata (1998). "Subadjunction of log canonical divisors. II". *Amer. J. Math.* 120.5, pp. 893–899. MR: 1646046 (cit. on p. 881).

Christer O. Kiselman (1978). "The partial Legendre transformation for plurisubharmonic functions". *Invent. Math.* 49.2, pp. 137–148. MR: 511187 (cit. on p. 879).

László Lempert and Liz Vivas (2013). "Geodesics in the space of Kähler metrics". *Duke Math. J.* 162.7, pp. 1369–1381. MR: 3079251 (cit. on p. 891).

Kefeng Liu, Xiaofeng Sun, and Shing-Tung Yau (2009). "Recent development on the geometry of the Teichmüller and moduli spaces of Riemann surfaces". In: *Surveys in differential geometry. Vol. XIV. Geometry of Riemann surfaces and their moduli spaces*. Vol. 14. Surv. Differ. Geom. Int. Press, Somerville, MA, pp. 221–259. MR: 2655329 (cit. on p. 896).

Zhiqin Lu (2001). "On the Hodge metric of the universal deformation space of Calabi-Yau threefolds". *J. Geom. Anal.* 11.1, pp. 103–118. MR: 1829350 (cit. on p. 897).

Toshiki Mabuchi (1986). "K-energy maps integrating Futaki invariants". *Tohoku Math. J. (2)* 38.4, pp. 575–593. MR: 867064 (cit. on pp. 893, 894).

– (1987). "Some symplectic geometry on compact Kähler manifolds. I". *Osaka J. Math.* 24.2, pp. 227–252. MR: 909015 (cit. on pp. 890, 891).

Fumio Maitani and Hiroshi Yamaguchi (2004). "Variation of Bergman metrics on Riemann surfaces". *Math. Ann.* 330.3, pp. 477–489. MR: 2099190 (cit. on pp. 883, 888).

J. Moser (1970/71). "A sharp form of an inequality by N. Trudinger". *Indiana Univ. Math. J.* 20, pp. 1077–1092. MR: 0301504 (cit. on p. 893).

Philipp Naumann (n.d.). "Curvature of higher direct images". arXiv: 1611.09117 (cit. on p. 897).

Takeo Ohsawa and Kenshō Takegoshi (1987). "On the extension of L^2 holomorphic functions". *Math. Z.* 195.2, pp. 197–204. MR: 892051 (cit. on p. 888).

E. Onofri (1982). "On the positivity of the effective action in a theory of random surfaces". *Comm. Math. Phys.* 86.3, pp. 321–326. MR: 677001 (cit. on p. 893).

Mihai Paun (n.d.). "Relative adjoint transcendental classes and Albanese maps of compact Kaehler manifolds with nef Ricci curvature". Advanced Studies in Pure Mathematics, Math. Soc of Japan, to appear. arXiv: 1209.2195 (cit. on p. 896).

Mihai Păun and Shigeharu Takayama (n.d.). "Positivity of twisted relative pluricanonical bundles and their direct images". arXiv: 1409.5504 (cit. on p. 898).

D. H. Phong and Jacob Sturm (2009). "Lectures on stability and constant scalar curvature". In: *Current developments in mathematics, 2007*. Int. Press, Somerville, MA, pp. 101–176. MR: 2532997 (cit. on p. 891).

András Prékopa (1973). "On logarithmic concave measures and functions". *Acta Sci. Math. (Szeged)* 34, pp. 335–343. MR: 0404557 (cit. on p. 877).

Hossein Raufi (n.d.). "Log concavity for matrix-valued functions and a matrix-valued Prékopa theorem". arXiv: 1311.7343 (cit. on p. 880).

– (2015). "Singular hermitian metrics on holomorphic vector bundles". *Ark. Mat.* 53.2, pp. 359–382. MR: 3391176 (cit. on p. 898).

Julius Ross and David Witt Nyström (n.d.). "On the Maximal Rank Problem for the Complex Homogeneous Monge-Ampére Equation". arXiv: 1610.02280 (cit. on p. 891).

– (2015). "Harmonic discs of solutions to the complex homogeneous Monge-Ampère equation". *Publ. Math. Inst. Hautes Études Sci.* 122, pp. 315–335. MR: 3415070 (cit. on p. 891).

H. L. Royden (1975). "Intrinsic metrics on Teichmüller space", pp. 217–221. MR: 0447636 (cit. on p. 896).

Georg Schumacher (2012). "Positivity of relative canonical bundles and applications". *Invent. Math.* 190.1, pp. 1–56. MR: 2969273 (cit. on pp. 886, 895, 897).

Stephen Semmes (1988). "Interpolation of Banach spaces, differential geometry and differential equations". *Rev. Mat. Iberoamericana* 4.1, pp. 155–176. MR: 1009123 (cit. on pp. 890, 891).

Yum Tong Siu (1986). "Curvature of the Weil-Petersson metric in the moduli space of compact Kähler-Einstein manifolds of negative first Chern class". In: *Contributions to several complex variables*. Aspects Math., E9. Friedr. Vieweg, Braunschweig, pp. 261–298. MR: 859202 (cit. on pp. 886, 897).

Wing-Keung To and Sai-Kee Yeung (2015). "Finsler metrics and Kobayashi hyperbolicity of the moduli spaces of canonically polarized manifolds". *Ann. of Math. (2)* 181.2, pp. 547–586. MR: 3275846 (cit. on p. 897).

Hajime Tsuji (2010). "Dynamical construction of Kähler-Einstein metrics". *Nagoya Math. J.* 199, pp. 107–122. MR: 2730413 (cit. on p. 895).

Eckart Viehweg (1995). *Quasi-projective moduli for polarized manifolds*. Vol. 30. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin, pp. viii+320. MR: 1368632 (cit. on p. 898).

Xu Wang (n.d.). "Curvature restrictions on a manifold with a flat Higgs bundle". arXiv: 1608.00777 (cit. on p. 897).

– (2017). "A curvature formula associated to a family of pseudoconvex domains". *Ann. Inst. Fourier (Grenoble)* 67.1, pp. 269–313. MR: 3612332 (cit. on p. 884).

Scott A. Wolpert (1986). "Chern forms and the Riemann tensor for the moduli space of curves". *Invent. Math.* 85.1, pp. 119–145. MR: 842050 (cit. on p. 896).

Shing Tung Yau (1978). "On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I". *Comm. Pure Appl. Math.* 31.3, pp. 339–411. MR: 480350 (cit. on p. 895).

Received 2017-12-06.

Bo Berndtsson

Department of Mathematics
Chalmers University of Technology
S-412 96 Göteborg
Sweden
bob@chalmers.se

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (903–936)

CANNON–THURSTON MAPS

Mahan Mj

Abstract

We give an overview of the theory of Cannon-Thurston maps which forms one of the links between the complex analytic and hyperbolic geometric study of Kleinian groups. We also briefly sketch connections to hyperbolic subgroups of hyperbolic groups and end with some open questions.

1 Kleinian groups and limit sets

The Lie group $PSL_2(\mathbb{C})$ can be viewed from three different points of view:

1. As a Lie group or a matrix group (**group-theoretic**).
2. As the isometry group of hyperbolic 3-space $\mathbf{H}^3$–the upper half space $\{(x, y, z) : z > 0\}$ equipped with the metric $ds^2 = \frac{dx^2+dy^2+dz^2}{z^2}$, or equivalently the open ball $\{(x, y, z) : (x^2 + y^2 + z^2) = r^2 < 1\}$ equipped with the metric $ds^2 = \frac{4(dx^2+dy^2+dz^2)}{(1-r^2)^2}$ (**geometric**).
3. As the group of Möbius transformations of the Riemann sphere $\hat{\mathbb{C}}$ (**complex dynamic/analytic**).

A finitely generated discrete subgroup $\Gamma \subset PSL_2(\mathbb{C})$ is called a *Kleinian group*. Depending on how we decide to look at $PSL_2(\mathbb{C})$, the group Γ can accordingly be thought of as a discrete subgroup of a Lie group, as the fundamental group of the complete hyperbolic 3-manifold $\mathbf{H}^3/\Gamma$, or in terms of its action by holomorphic automorphisms of $\hat{\mathbb{C}}$. If Γ is not virtually abelian, it is called *non-elementary*. Henceforth, unless explicitly stated otherwise, we shall assume that *all Kleinian groups in this article are non-elementary*. If Γ can be conjugated by an element of $PSL_2(\mathbb{C})$ to be contained in $PSL_2(\mathbb{R})$ it is referred

Research partly supported by a DST JC Bose Fellowship.
MSC2010: primary 57M50; secondary 30F40, 20F65, 20F67.

to as a *Fuchsian group*[1]. If Γ is abstractly isomorphic to $\pi_1(S)$, the fundamental group of a closed surface S, we shall refer to it as a *surface Kleinian group*.

In the 1960's Ahlfors and Bers studied deformations of Fuchsian surface groups in $PSL_2(\mathbb{C})$, giving rise to the theory of quasi-Fuchsian groups. Their techniques were largely complex analytic in nature. In the 1970's and 80's, the field was revolutionized by Thurston, who brought in a rich and varied set of techniques from three-dimensional hyperbolic geometry. A conjectural picture of the deep relationships between the analytic and geometric points of view was outlined by Thurston in his visionary paper Thurston [1982]. Perhaps the most well-known problem (predating Thurston) in this line of study was the Ahlfors' measure zero conjecture, resolved in the last decade by Brock-Canary-Minsky Brock, Canary, and Y. N. Minsky [2012]. Another (more topological) well-known problem that also predates Thurston asks if limit sets of Kleinian groups are locally connected. This is the specific problem that will concern us here. We need to fix some terminology and notation first. Identify the Riemann sphere $\hat{\mathbb{C}}$ with the sphere at infinity S^2 of $\mathbf{H}^3$. Thus, S^2 encodes the 'ideal' boundary of $\mathbb{H}^3$, consisting of asymptote classes of geodesics. By adjoining S^2 to $\mathbf{H}^3$, we obtain the closed 3-ball $\mathbf{D}^3$. The topology on S^2 is the usual one induced by the round metric given by the angle subtended at the origin $0 \in \mathbf{D}^3$. The geodesics turn out to be semicircles meeting the boundary S^2 at right angles.

Definition 1.1. *The* limit set Λ_Γ *of the Kleinian group* Γ *is the collection of accumulation points of a* Γ*-orbit* $\Gamma \cdot z$ *for some (any)* $z \in \hat{\mathbb{C}}$.

The limit set Λ_Γ is independent of z and may be regarded as the locus of chaotic dynamics of Γ on $\hat{\mathbb{C}}$. For non-elementary Γ and any $z \in \Lambda_\Gamma$, $\Gamma \cdot z$ is dense in Λ_Γ. Hence Λ_Γ is the smallest closed non-empty $\Gamma-$invariant subset of $\hat{\mathbb{C}}$. If we take $z \in \mathbf{H}^3$ instead, then the collection of accumulation points of any Γ-orbit $\Gamma \cdot z \subset \mathbf{D}^3$ is also Λ_Γ.

Definition 1.2. *The complement of the limit set* $\hat{\mathbb{C}} \setminus \Lambda_\Gamma$ *is called the domain of discontinuity* Ω_Γ *of* Γ.

The Kleinian group Γ acts freely and properly discontinuously on Ω_Γ and Ω_Γ/Γ is a (possibly disconnected) Riemann surface.

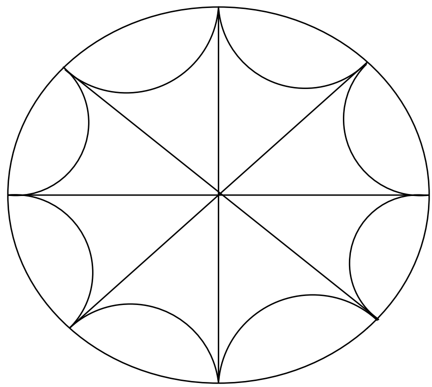
1.1 Fuchsian and Quasi-Fuchsian Groups.

We first give an explicit example of a Fuchsian group.

An example of a Fuchsian group: Consider the standard identification space description of the genus two orientable surface Σ_2 as an octagon with edge labels $a_1, b_1, a_1^{-1}, b_1^{-1}, a_2, b_2, a_2^{-1}, b_2^{-1}$. A hyperbolic metric on Σ_2 is one where each point has a small neighborhood isometric to a small disk in $\mathbf{H}^2$. By the Poincaré polygon theorem, it suffices to find a

[1]Both Fuchsian and Kleinian groups were discovered by Poincaré.

regular hyperbolic octagon (with all sides equal and all angles equal) with each interior angle equal to $\frac{2\pi}{8}$. Now, the infinitesimal regular octagon at the tangent space to the origin has interior angles equal to $\frac{3\pi}{4}$. Also the ideal regular octagon in $\mathbb{H}^2$ has all interior angles zero. See figure below.



By the Intermediate Value Theorem, there exists an intermediate regular octagon with all interior angles equal to $\frac{\pi}{4}$. The group G that results from side-pairing transformations corresponds to a Fuchsian group, or equivalently, a discrete faithful representation ρ of $\pi_1(\Sigma_2)$ into $PSL_2(\mathbb{R})$. Equivalently we may regard ρ as a representation of $\pi_1(\Sigma_2)$ into $PSL_2(\mathbb{C})$ with image that can be conjugated to lie in $PSL_2(\mathbb{R})$. Alternately, $\rho(\pi_1(\Sigma_2))$ preserves a totally geodesic plane in $\mathbf{H}^3$. The limit set of $G = \rho(\pi_1(\Sigma_2))$ is then a round circle.

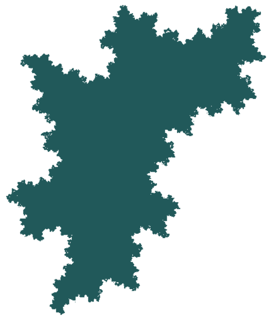
Quasi-Fuchsian groups: If we require the limit set to be only topologically a circle, i.e. a Jordan curve, then we obtain a more general class of Kleinian groups:

Definition 1.3. *Let $\rho : \pi_1(S) \to PSL_2(\mathbb{C})$ be a discrete faithful representation such that the limit set of $G = \rho(\pi_1(S))$ is a Jordan curve in S^2. Then G is said to be* quasi-Fuchsian. *The collection of conjugacy classes of quasi-Fuchsian with the complex analytic structure inherited from $PSL_2(\mathbb{C})$ is denoted as $QF(S)$.*

The domain of discontinuity Ω of a quasi-Fuchsian G consists of two open invariant topological disks Ω_1, Ω_2 in $\hat{\mathbb{C}}$. Hence the quotient Ω/G is the disjoint union $\Omega_1/G \sqcup \Omega_2/G$ and we have a map $\tau : QF(S) \to Teich(S) \times Teich(S)$, where $Teich(S)$ denotes the Teichmüller space of S–the space of marked hyperbolic (or complex) structures on S. The *Bers simultaneous Uniformization Theorem* asserts:

Theorem 1.4. *$\tau : QF(S) \to Teich(S) \times Teich(S)$ is bi-holomorphic.*

Thus, given any two conformal structures T_1, T_2 on a surface, there is a unique discrete quasi-Fuchsian G (up to conjugacy) whose limit set Λ_G is topologically a circle, and the quotient of whose domain of discontinuity is $T_1 \sqcup T_2$. See figure below Kabaya [2016], where the inside and the outside of the Jordan curve correspond to Ω_1, Ω_2.



1.2 Degenerate groups and the Ending Lamination Theorem. Quasi-Fuchsian were studied by Ahlfors and Bers analytically as deformations of Fuchsian groups. Thurston [1980] introduced a new set of geometric techniques in the study.

Definition 1.5. *The* convex hull CH_G *of* Λ_G *is the smallest non-empty closed convex subset of* $\mathbb{H}^3$ *invariant under* G.

Let $M = \mathbf{H}^3/G$. *The quotient of* CH_G *by* G *is called the convex core* $CC(M)$ *of* M.

The convex hull CH_G can be constructed by joining all distinct pairs of points on Λ_G by bi-infinite geodesics, iterating this construction and finally taking the closure. It can also be described as the closure of the union of all ideal tetrahedra, whose vertices lie in Λ_G. The convex core $CC(M)$ is homeomorphic to $S \times [0,1]$.

The distance between the boundary components $S \times \{0\}$ and $S \times \{1\}$ in the convex core $CC(M)$, measured with respect to the hyperbolic metric, is a crude geometric measure of the complexity of the quasi Fuchsian group G. We shall call it the *thickness* t_G of $CC(M)$, or of the quasi Fuchsian group G. (We note here parenthetically that the notions of convex hull CH_G and convex core $CC(M)$ go through for any Kleinian group G and the associated complete hyperbolic manifold $\mathbf{H}^3/G$.) For quasi-Fuchsian groups, we ask:

Question 1.6. *What happens as thickness tends to infinity?*

To address this question more precisely we need to introduce a topology on the space of representations.

Definition 1.7. *A sequence of representations $\rho_n : \pi_1(S) \to PSL_2(\mathbb{C})$ is said to converge* algebraically *to $\rho_\infty : \pi_1(S) \to PSL_2(\mathbb{C})$ if for all $g \in \pi_1(S)$, $\rho_n(g) \to \rho_\infty(g)$ in $PSL_2(\mathbb{C})$. The collection of conjugacy classes of discrete faithful representations of $\pi_1(S)$ into $PSL_2(\mathbb{C})$ equipped with the algebraic topology is denoted as $AH(S)$.*

It is not even clear a priori that, as t_G tends to infinity, limits exist in $AH(S)$. However, Thurston's *Double Limit Theorem* Thurston [1986], M. Kapovich [2001], and Otal [2001] guarantees that if we have a sequence G_n with thickness t_{G_n} tending to infinity, subsequential limits (in the space of discrete faithful representations with a suitable topology, see Definition 5.7) do in fact exist.

Geodesic Laminations:

Definition 1.8. *A geodesic lamination on a hyperbolic surface is a foliation of a closed subset with geodesics, i.e. it is a closed set given as a disjoint union of geodesics (closed or bi-infinite).*

A geodesic lamination on a surface may further be equipped with a transverse (non-negative) measure to obtain a *measured lamination*. The space of measured (geodesic) laminations on S is then a positive cone in a vector space and is denoted as $\mathfrak{ML}(S)$. It can be projectivized to obtain the space of projectivized measured laminations $\mathfrak{PML}(S)$. It was shown by Thurston Fathi, Laudenbach, and Poenaru [1979] that $\mathfrak{PML}(S)$ is homeomorphic to a sphere and can be adjoined to $Teich(S)$ compactifying it to a closed ball.

Definition 1.9. *Thurston [1980, Definition 8.8.1] A* pleated surface *in a hyperbolic three-manifold N is a complete hyperbolic surface S of finite area, together with an isometric map $f : S \to N$ such that every $x \in S$ is in the interior of some geodesic segment (in S) which is mapped by f to a geodesic segment (in N). Also, f must take every cusp (corresponding to a maximal parabolic subgroup) of S to a cusp of N*

The pleating locus *of the pleated surface $f : S \to M$ is the set $\gamma \subset S$ consisting of those points in the pleated surface which are in the interior of unique geodesic segments mapped to geodesic segments.*

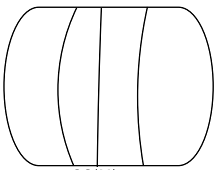
Proposition 1.10. *Thurston [ibid., Proposition 8.8.2] The pleating locus γ is a geodesic lamination on S. The map f is totally geodesic in the complement of γ.*

Thurston further shows Thurston [ibid., Ch. 8] that the boundary components of the convex core $CC(M)$ are pleated surfaces.

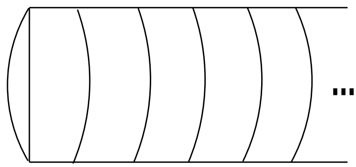
As thickness (Question 1.6) tends to infinity, the convex core may converge (up to extracting subsequences) to one of two kinds of convex manifolds $CC(M_\infty)$ (see schematic diagram below):

1. $CC(M_\infty)$ is homeomorphic to $S \times [0, \infty)$. Here $S \times \{0\}$ is the single boundary component of the convex core $CC(M_\infty)$. Such a manifold is called *simply degenerate* and the corresponding Kleinian group a simply degenerate surface Kleinian group. In this case the limit set Λ_G is a dendrite (topologically a tree) and the domain of discontinuity Ω_Γ is homeomorphic to an open disk with Ω_Γ/G a Riemann surface homeomorphic to $S \times \{0\}$. The end of M_∞ facing Ω_Γ/G is called a *geometrically finite end*.

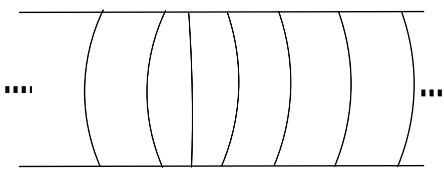
2. $CC(M_\infty)$ is homeomorphic to $S \times (-\infty, \infty)$. Such a manifold is called *doubly degenerate* and the corresponding Kleinian group a doubly degenerate surface Kleinian group. In this case the limit set Λ_G is all of S^2 and the domain of discontinuity Ω_Γ is empty.



CC(M) compact



CC(M) for simply degenerate



CC(M) for doubly degenerate

The ends (one for simply degenerate and two for doubly degenerate) of $CC(M_\infty)$ are called the *degenerate end(s)* of M. Thurston [1980] and Bonahon [1986] show that any such degenerate end E is *geometrically tame*, i.e. there is a sequence of simple closed curves $\{\sigma_n\}$ on S such that their geodesic realizations in $CC(M_\infty)$ exit E as n tends to ∞. Further Thurston [1980, Ch. 9], the limit of any such exiting sequence (in $\mathcal{PML}(S)$; the topology on $\mathcal{PML}(S)$ is quite close to the Hausdorff topology on S) is a unique lamination λ called the *ending lamination* of E. Thus a doubly degenerate manifold has two ending laminations, one for each degenerate end, while a simply degenerate manifold has a geometrically finite end corresponding to a Riemann surface $\Omega_\Gamma/G (\in Teich(S))$ at infinity and a degenerate end corresponding to an ending lamination.

These two pieces of information–Riemann surfaces at infinity and ending laminations–give the *end-invariants* of M. The ending lamination for a geometrically infinite end does not depend on the reference hyperbolic structure on S. It may be regarded as a purely topological piece of data associated to such an end.

We may thus think of the ending lamination as the analog, in the case of geometrically infinite ends, of the conformal structure at infinity for a geometrically infinite end.

The Ending Lamination Theorem of Brock-Canary-Minsky then takes the place of the Bers' simultaneous uniformization theorem and asserts:

Theorem 1.11. *Y. Minsky [2010] and Brock, Canary, and Y. N. Minsky [2012] Let M be a simply or doubly degenerate manifold. Then M is determined up to isometry by its end-invariants.*

Thus, the Ending Lamination Theorem justifies the following and may be considered an analog of Mostow Rigidity for infinite covolume Kleinian groups.

Slogan 1.12. *Topology implies Geometry.*

In order to complete the picture of the Ahlfors-Bers theory to study degenerate surface Kleinian groups G in terms of the dynamics of G on the Riemann sphere $\hat{\mathbb{C}}$, the following issue remains to be addressed:

Question 1.13. *Can the data of the ending lamination(s) be extracted from the dynamics of G on $\hat{\mathbb{C}}$?*

In more informal terms,

Question 1.14. *Is the geometric object "at infinity" of the quotient manifold $\mathbf{H}^3/\Gamma$ (i.e. the ending lamination) determined by the dynamics of Γ at infinity (i.e. the action of Γ on S^2)?*

We will make these questions precise below. The attempt to make Question 1.13 precise brings us to the following.

2 Cannon-Thurston maps

2.1 The main theorem for closed surface Kleinian groups.

In Thurston [1982, Problem 14], Thurston raised the following question, which is at the heart of the work we discuss here:

Question 2.1. *Suppose Γ has the property that $(\mathbf{H}^3 \cup \Omega_\Gamma)/\Gamma$ is compact. Then is it true that the limit set of any other Kleinian group Γ' isomorphic to Γ is the continuous image of the limit set of Γ, by a continuous map taking the fixed points of an(y) element γ to the fixed points of the corresponding element γ'?*

A special case of Question 2.1 was answered affirmatively in seminal work of Cannon and Thurston [1985, 2007]:

Theorem 2.2. *Cannon and Thurston [2007] Let M be a closed hyperbolic 3-manifold fibering over the circle with fiber Σ. Let $\widetilde{\Sigma}$ and $\widetilde{M}$ denote the universal covers of F and M respectively. After identifying $\widetilde{\Sigma}$ (resp. $\widetilde{M}$) with $\mathbf{H}^2$ (resp. $\mathbf{H}^3$), we obtain the compactification $\mathbb{D}^2 = \mathbf{H}^2 \cup S^1$ (resp. $\mathbb{D}^3 = \mathbf{H}^3 \cup S^2$) by attaching the circle S^1 (resp. the sphere S^2) at infinity. Let $i : \Sigma \to M$ denote the inclusion map of the fiber and $\widetilde{i} : \widetilde{\Sigma} \to \widetilde{M}$ the lift to the universal cover. Then $\widetilde{i}$ extends to a continuous map $\hat{i} : \mathbb{D}^2 \to \mathbb{D}^3$.*

An amazing implication of Theorem 2.2 is that $\widetilde{\Sigma}$ is an embedded disk in (the ball model) of $\mathbf{H}^3$ such that its boundary on the sphere S^2 is space-filling! See the following diagram by Thurston Thurston [1982, Figure 10] that illustrates 'a pattern of identification of a circle, here represented as the equator, whose quotient is topologically a sphere. This defines, topologically a sphere-filling curve.'

A version of Question 2.1 was raised by Cannon and Thurston in the context of closed surface Kleinian groups:

Question 2.3. *Cannon and Thurston [2007, Section 6] Suppose that a closed surface group $\pi_1(S)$ acts freely and properly discontinuously on $\mathbb{H}^3$ by isometries such that the quotient manifold has no accidental parabolics (here this just means that the image of $\pi_1(S)$ in $PSL_2(\mathbb{C})$ has no parabolics). Does the inclusion $\tilde{i} : \widetilde{S} \to \mathbb{H}^3$ extend continuously to the boundary?*

Continuous boundary extensions as in Question 2.3, if they exist, are called *Cannon-Thurston maps*. Question 2.3 is intimately related to a much older question c.f. Abikoff [1976] asking if limit sets are locally connected:

Question 2.4. *Let Γ be a finitely generated Kleinian group such that the limit set Λ_Γ is connected. Is Λ_Γ locally connected?*

It is shown in Cannon and Thurston [2007] that for simply degenerate surface Kleinian groups, Questions 2.3 and 2.4 are equivalent, via the Caratheodory extension Theorem.

The following Theorem of Mj [2014a] Mj [2014b] answers questions 2.3 and 2.4 affirmatively:

Theorem 2.5. *Let $\rho(\pi_1(S)) = G \subset PSL_2(\mathbb{C})$ be a simply or doubly degenerate (closed) surface Kleinian group. Let $M = \mathbf{H}^3/G$ and $i : S \to M$ be an embedding inducing a homotopy equivalence. Let $\widetilde{i} : \widetilde{S} \to \mathbf{H}^3$ denote a lift of i between universal covers. Let $\mathbb{D}^2, \mathbb{D}^3$ denote the compactifications. Then a Cannon-Thurston map $\hat{i} : \mathbb{D}^2 \to \mathbb{D}^3$ exists.*

Let $\partial i : S^1 \to S^2$ denote the restriction of $\hat{i}$ to the ideal boundaries. Then for $p \neq q$, $\partial i(p) = \partial i(q)$ if and only if p, q are the ideal end-points of a leaf of an ending lamination or ideal end-points of a complementary ideal polygon of an ending lamination.

The second part of Theorem 2.5 shows that the data of the ending lamination can be recovered from the Cannon-Thurston map and so we have an affirmative answer to Question 1.13. In conjunction with the Ending Lamination Theorem 1.11, this establishes the slogan:

Slogan 2.6. *Dynamics on the limit set determines geometry in the interior.*

A number of authors have contributed to the resolution of the above questions. Initially it was believed Abikoff [1976] that Question 2.4 had a negative answer for simply degenerate Kleinian groups. Floyd [1980] proved the corresponding theorem for geometrically finite Kleinian groups. Then in the early 80's Cannon and Thurston [1985] proved Theorem 2.2. This was extended by Y. N. Minsky [1994], Klarreich [1999], Alperin, Dicks, and Porti [1999], B. H. Bowditch [2013] and B. H. Bowditch [2007], McMullen [2001], Miyachi [2002] and the author 1998; 2010; 2009-2010; 2016 for various special cases. The general surface group case was accomplished in Mj [2014a] and the general Kleinian group case in Mj [2017a].

2.2 Geometric Group Theory. We now turn to a generalization of Question 2.3 to a far more general context. After the introduction of hyperbolic metric spaces by Gromov [1987], Question 2.3 was extended by the author Mitra [1997b], Bestvina [2004], and Mitra [1998c] to the context of a hyperbolic group H acting freely and properly discontinuously by isometries on a hyperbolic metric space X. Any hyperbolic X has a (Gromov) boundary ∂X given by asymptote-classes of geodesics. Adjoining ∂X to X we get the Gromov compactification $\widehat{X}$.

There is a natural map $i : \Gamma_H \to X$, sending vertices of Γ_H to the $H-$orbit of a point $x \in X$, and connecting images of adjacent vertices by geodesic segments in X. Let $\widehat{\Gamma_H}, \widehat{X}$ denote the Gromov compactification of Γ_H, X respectively. The analog of Question 2.3 is the following:

Question 2.7. *Does $i : \Gamma_H \to X$ extend continuously to a map $\hat{i} : \widehat{\Gamma_H} \to \widehat{X}$?*

Continuous extensions as in Question 2.7 are also referred to as *Cannon-Thurston maps* and make sense when Γ_H is replaced by an arbitrary hyperbolic metric space Y. A simple and basic criterion for the existence of Cannon-Thurston maps was established in Mitra [1998a,b]:

Lemma 2.8. *Let $i : (Y, y) \to (X, x)$ be a proper map between (based) Gromov-hyperbolic spaces. A continuous extension (also called a Cannon-Thurston map) $\hat{i} : \widehat{Y} \to \widehat{X}$ exists if and only if the following holds:*
There exists a non-negative proper function $M : \mathbb{N} \to \mathbb{N}$, such that if $\lambda = [a, b]_Y$ is a geodesic lying outside an N-ball around y, then any geodesic segment $[i(a), i(b)]_X$ in X joining $i(a), i(b)$ lies outside the $M(N)$-ball around $x = i(y)$.

In the generality above Question 2.7 turns out to have a negative answer. An explicit counterexample to Question 2.7 was recently found by Baker and T. R. Riley [2013] in the context of small cancellation theory. The counterexample uses Lemma 2.8 to rule out the existence of Cannon-Thurston maps. Further, Matsuda and Oguni [2014] further developed Baker and Riley's counterexample to show that given a(ny) non-elementary hyperbolic group H, there exists hyperbolic group G such that $H \subset G$ and there is no Cannon-Thurston map for the inclusion. We shall furnish positive answers to Question 2.7 in a number of special cases in Section 6.

3 Closed 3-manifolds

3.1 3-manifolds fibering over the circle. We start by giving a sketch of the proof of Theorem 2.2, The proof is coarse-geometric in nature and follows Mitra [1998a,b]. We recall a couple of basic Lemmata we shall be needing from Mitra [1998b]. The following says that nearest point projection onto a geodesic in a hyperbolic space is coarsely Lipschitz.

Lemma 3.1. *Let (X, d) be a δ-hyperbolic metric space. Then there exists a constant $C \geq 1$ such that the following holds:*
Let $\lambda \subset X$ be a geodesic segment and let $\Pi : X \to \lambda$ be a nearest point projection. Then $d(\Pi(x), \Pi(y)) \leq C d(x, y)$ for all $x, y \in X$.

The next Lemma says that nearest point projections and quasi-isometries almost commute.

Lemma 3.2. *Let (X, d) be a δ-hyperbolic metric space. Given (K, ϵ), there exists C such that the following holds:*

Let $\lambda = [a, b]$ be a geodesic segment in X. Let $p \in X$ be arbitrary and let q be a nearest point projection of p onto λ. Let ϕ be a $(K, \epsilon)-$quasi-isometry from X to itself and let $\Phi(\lambda) = [\phi(a), \phi(b)]$ be a geodesic segment in X joining $\phi(a), \phi(b)$. Let r be a nearest point projection of $\phi(p)$ onto $\Phi(\lambda)$. Then $d(r, \phi(q)) \leq C$.

Sketch of Proof: The proof of Lemma 3.2 follows from the fact that a geodesic tripod T (built from $[a, b]$ and $[p, q]$) is quasiconvex in hyperbolic space and that a quasi-isometric image $\phi(T)$ of T lies close to a geodesic tripod T' built from $[\phi(a), \phi(b)]$ and $[\phi(p), r]$. Hence the image $\phi(q)$ of the centroid q of T lies close to the centroid r of T'. □

3.2 The key tool: hyperbolic ladder. The key idea behind the proof of Theorem 2.2 and its generalizations in Mitra [1998a,b] is the construction of a *hyperbolic ladder* $\mathfrak{L}_\lambda \subset \widetilde{M}$ for *any* geodesic in $\widetilde{\Sigma}$. The universal cover $\widetilde{M}$ fibers over $\mathbb{R}$ with fibers $\widetilde{\Sigma}$. Since the context is geometric group theory, we discretize this as follows. Replace $\widetilde{\Sigma}$ and $\widetilde{M}$ by quasi-isometric models in the form of Cayley graphs $\Gamma_{\pi_1(\Sigma)}$ and $\Gamma_{\pi_1(M)}$ respectively. Let us denote $\Gamma_{\pi_1(\Sigma)}$ by Y and $\Gamma_{\pi_1(M)}$ by X. The projection of $\widetilde{M}$ to the base $\mathbb{R}$ is discretized accordingly giving a model that can be thought of as (and is quasi-isometric to) a *tree T of spaces*, where

1. T is the simplicial tree underlying $\mathbb{R}$ with vertices at $\mathbb{Z}$.
2. All the vertex and edge spaces are (intrinsically) isometric to Y.
3. The edge space to vertex space inclusions are qi-embeddings.

We summarize this by saying that X is a tree T of spaces satisfying the *qi-embedded condition* Bestvina and Feighn [1992].

Given a geodesic segment $[a, b] = \lambda = \lambda_0 \subset Y$, we now sketch the promised construction of the *ladder* $\mathfrak{L}_\lambda \subset X$ containing λ. Index the vertices by $n \in \mathbb{Z}$. Since the edge-to-vertex space inclusions are quasi-isometries, this induces a quasi-isometry ϕ_n from the vertex space Y_n to the vertex space Y_{n+1} for $n \geq 0$. A similar quasi-isometry ϕ_{-n} exists from Y_{-n} to the vertex space $Y_{-(n+1)}$. These quasi-isometries are defined on the vertex sets of Y_n, $n \in \mathbb{Z}$. ϕ_n induces a map Φ_n from geodesic segments in Y_n to geodesic segments in Y_{n+1} for $n \geq 0$ by sending a geodesic in Y_n joining a, b to a geodesic in Y_{n+1} joining $\phi_n(a), \phi_n(b)$. Similarly, for $n \leq 0$. Inductively define:

- $\lambda_{j+1} = \Phi_j(\lambda_j)$ for $j \geq 0$,
- $\lambda_{-j-1} = \Phi_{-j}(\lambda_{-j})$ for $j \geq 0$,
- $\mathfrak{L}_\lambda = \bigcup_j \lambda_j$.

$\mathfrak{L}_\lambda$ turns out to be quasiconvex in X. To prove this, we construct a coarsely Lipschitz retraction $\Pi_\lambda : \bigcup_j Y_j \to \mathfrak{L}_\lambda$ as follows.

On Y_j define $\pi_j(y)$ to be a nearest-point projection of y onto λ_j and define

$$\Pi_\lambda(y) = \pi_j(y),\ for\ y \in Y_j.$$

The following theorem asserts that Π_λ is coarsely Lipschitz.

Theorem 3.3. *Mitra [1998a,b] and Mj [2010] There exists $C \geq 1$ such that for any geodesic $\lambda \subset Y$,*

$$d_X(\Pi_\lambda(x), \Pi_\lambda(y)) \leq C d_X(x, y)$$

for $x, y \in \bigcup_i Y_i$.

Sketch of Proof:
The proof requires only the hyperbolicity of Y, but not that of X. It suffices to show that for $d_X(x, y) = 1$, $d_X(\Pi_\lambda(x), \Pi_\lambda(y)) \leq C$. Thus x, y may be thought of as

1. either lying in the same Y_j. This case follows directly from Lemma 3.1.

2. or lying vertically one just above the other. Then (up to a bounded amount of error), we can assume without loss of generality, that $y = \phi_j(x)$. This case now follows from Lemma 3.2.

Since a coarse Lipschitz retract of a hyperbolic metric space is quasiconvex, we immediately have:

Corollary 3.4. *If (X, d_X) is hyperbolic, there exists $C \geq 1$ such that for any λ, $\mathfrak{L}_\lambda$ is $C-$quasiconvex.*

Note here that we have not used any feature of Y except its hyperbolicity. In particular, we do not need the specific condition that $Y = \widetilde{\Sigma}$. We are now in a position to prove a generalization of Theorem 2.2.

Theorem 3.5. *Mj [2010] Let (X, d) be a hyperbolic tree (T) of hyperbolic metric spaces satisfying the qi-embedded condition, where T is $\mathbb{R}$ or $[0, \infty)$ with vertex and edge sets Y_j as above, $j \in \mathbb{Z}$. Assume (as above) that the edge-to-vertex inclusions are quasi-isometries. For $i : Y_0 \to X$ there is a Cannon-Thurston map $\hat{i} : \widehat{Y_0} \to \widehat{X}$.*

Proof. Fix a basepoint $y_0 \in Y_0$. By Lemma 2.8 and quasiconvexity of $\mathfrak{L}_\lambda$ (Corollary 3.4), it suffices to show that for all $M \geq 0$ there exists $N \geq 0$ such that if a geodesic segment λ lies outside the N-ball about $y_0 \in Y_0$, then $\mathfrak{L}_\lambda$ lies outside the M-ball around $y_0 \in X$. Equivalently, we need a proper function $M(N) : \mathbb{N} \to \mathbb{N}$.

Since Y_0 is properly embedded in X, there exists a proper function $g : \mathbb{N} \to \mathbb{N}$ such that λ lies outside the $g(N)$-ball about $y_0 \in X$.

Let p be any point on $\mathfrak{L}_\lambda$. Then $p = p_j \in Y_j$ for some j. Assume without loss of generality that $j \geq 0$. It is not hard to see that there exists C_0, depending only on X, such that for any such p_j, there exists $p_{j-1} \in Y_{j-1}$ with $d(p_j, p_{j-1}) \leq C_0$. It follows inductively that there exists $y \in \lambda = \lambda_0$ such that $d_X(y, p) \leq C_0 j$. Hence, by the triangle inequality, $d_X(y_0, p) \geq g(N) - C_0 j$.

Next, looking at the 'vertical direction', $d_X(y_0, p) \geq j$ and hence

$$d_X(y_0, p) \geq \max(g(N) - C_0 j, j) \geq \frac{g(N)}{A+1}.$$

Defining $M(N) = \frac{g(N)}{A+1}$, we see that $M(N)$ is a proper function of N and we are done. □

3.3 Quasiconvexity. The structure of Cannon-Thurston maps in Section 3.1 can be used to establish quasiconvexity of certain subgroups of $\pi_1(M)$. Let $H \subset \pi_1(\Sigma)$ be a finitely generated infinite index subgroup of the fiber group. Then, due to the LERF property for surface subgroups, a Theorem of Scott P. Scott [1978], there is a finite sheeted cover where H is geometric, i.e. it is carried by a proper embedded subsurface of (a finite sheeted cover of) Σ. But such a proper subsurface cannot carry a leaf of the stable or unstable foliations $\mathfrak{F}_s$ or $\mathfrak{F}_u$. This gives us the following Theorem of Scott and Swarup:

Theorem 3.6. *G. P. Scott and Swarup [1990] Let M be a closed hyperbolic 3-manifold fibering over the circle with fiber Σ. Let $H \subset \pi_1(\Sigma)$ be a finitely generated infinite index subgroup of the fiber group in $\pi_1(M)$. Then H is quasiconvex in $\pi_1(M)$.*

Theorem 3.6 has been generalized considerably to the context of convex cocompact subgroups of the mapping class group and $Out(F_n)$ by a number of authors Dowdall, Kent, and C. J. Leininger [2014], Dowdall, I. Kapovich, and Taylor [2016], Dowdall and Taylor [2018, 2017], and Mj and Rafi [2015].

4 Kleinian surface groups: Model Geometries

In this section we shall describe a sequence of models for degenerate ends of 3-manifolds following Y. N. Minsky [2001, 1994] and Mj [2010, 2009-2010, 2016] and Y. Minsky [2010], Brock, Canary, and Y. N. Minsky [2012], and Mj [2014a] and indicate how to generalize the ladder construction of Section 3.2 incorporating electric geometry Farb [1998]. Let X be a hyperbolic metric space, e.g. $\mathbf{H}^3$. Let $\mathcal{H}_X$ be a collection of disjoint convex subsets. Roughly speaking, electrification equips each element of $\mathcal{H}_X$ with the zero metric, while preserving the metric on $X \setminus (\bigcup_{H \in \mathcal{H}_X} H)$. The resulting electrified space $\mathcal{E}(X, \mathcal{H}_X)$ is still Gromov hyperbolic under extremely mild conditions and hyperbolic geodesics in X can be recovered from electric geodesics in the electrified space $\mathcal{E}(X, \mathcal{H}_X)$. This will allow us to establish the existence of Cannon-Thurston maps. We shall focus on closed surfaces and follow the summary in Lecuire and Mj [2016] for the exposition.

The topology of each building block is simple: it is homeomorphic to $S \times [0, 1]$, where S is a closed surface of genus greater than one. Geometrically, the top and bottom boundary components in the first three model geometries are uniformly bi-Lipschitz to a fixed hyperbolic structure on S. Assume therefore that S is equipped with such a fixed hyperbolic structure. We do so henceforth. The different types of geometries of the blocks make for different model geometries of ends.

Definition 4.1. *A model E_m is said to be built up of blocks of some prescribed geometries* glued end to end*, if*

1. *E_m is homeomorphic to $S \times [0, \infty)$*
2. *There exists $L \geq 1$ such that $S \times [i, i+1]$ is $L-$bi-Lipschitz to a block of one of the three prescribed geometries: bounded, i-bounded or amalgamated (see below).*

*$S \times [i, i+1]$ will be called the $(i+1)$*th block *of the model E_m.*

The thickness *of the $(i+1)$th block is the length of the shortest path between $S \times \{i\}$ and $S \times \{i+1\}$ in $S \times [i, i+1] (\subset E_m)$.*

4.1 Bounded geometry.

Y. N. Minsky [2001, 1994] calls an end E of a hyperbolic 3-manifold to be of bounded geometry if there are no arbitrarily short closed geodesics in E.

Definition 4.2. *Let $B_0 = S \times [0, 1]$ be given the product metric. If B is $L-$bi-Lipschitz homeomorphic to B_0, it is called an $L-$*thick block.

An end E is said to have a model of bounded geometry if there exists L such that E is bi-Lipschitz homeomorphic to a model manifold E_m consisting of gluing $L-$thick blocks end-to-end.

It follows from work of of Minsky Y. N. Minsky [1993] that if E is of bounded geometry, it has a model of bounded geometry. The existence of Cannon-Thurston maps in this setup is then a replica of the proof of Theorem 3.5.

4.2 i-bounded Geometry.

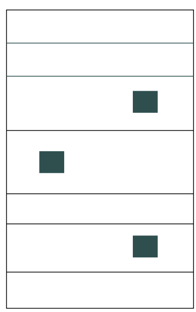
Definition 4.3. *Mj [2009-2010] An end E of a hyperbolic 3-manifold M has* i-bounded geometry *if the boundary torus of every Margulis tube in E has bounded diameter.*

We give an alternate description. Fix a closed hyperbolic surface S. Let $\mathcal{C}$ be a finite collection of (not necessarily disjoint) simple closed geodesics on S. Let $N_\epsilon(\sigma_i)$ denote an ϵ neighborhood of σ_i, $\sigma_i \in \mathcal{C}$, were ϵ is small enough to ensure that lifts of $N_\epsilon(\sigma_i)$ to $\widetilde{S}$ are disjoint.

Definition 4.4. *Let $I = [0,3]$. Equip $S \times I$ with the product metric. Let $B^c = (S \times I - \cup_j N_\epsilon(\sigma_j) \times [1,2])$, equipped with the induced path-metric. Here $\{\sigma_j\}$ is a subcollection of $\mathcal{C}$ consisting of disjoint curves. Perform Dehn filling on some $(1,n)$ curve on each resultant torus component of the boundary of B^c (the integers n are quite arbitrary and may vary: we omit subscripts for expository ease). We call n the* twist coefficient. *Foliate the relevant torus boundary component of B^c by translates of $(1,n)$ curves. Glue in a solid torus Θ, which we refer to as a* Margulis tube, *with a hyperbolic metric foliated by totally geodesic disks bounding the $(1,n)$ curves.*

The resulting copy of $S \times I$ thus obtained, equipped with the above metric is called a thin block.

Definition 4.5. *A model manifold E_m of* i-bounded geometry *is built out of gluing $L-$thick and thin blocks end-to-end (for some L) (see schematic diagram below where the black squares indicate Margulis tubes and the horizontal rectangles indicate the blocks).*



It follows from work in Mj [ibid.] that

Proposition 4.6. *An end E of a hyperbolic 3-manifold M has* i-bounded geometry *if and only if it is bi-Lipschitz homeomorphic to a model manifold E_m of i-bounded geometry.*

We give a brief indication of the construction of $\mathcal{L}_\lambda$ and the proof of the existence of Cannon-Thurston maps in this case. First electrify all the Margulis tubes, i.e. equip them with a zero metric (see Farb [1998] for details on relative hyperbolicity and electric geometry). This ensures that in the resulting electric geometry, each block is of bounded geometry. More precisely, there is a (metric) product structure on $S \times [0,3]$ such that each $\{x\} \times [0,3]$ has uniformly bounded length in the electric metric.

Further, since the curves in $\mathcal{C}$ are electrified in a block, Dehn twists are isometries from $S \times \{1\}$ to $S \times \{2\}$ in a thin block. This allows the construction of $\mathcal{L}_\lambda$ to go through as before and ensures that it is quasiconvex in the resulting electric metric.

Finally given an electric geodesic lying outside large balls modulo Margulis tubes one can recover a genuine hyperbolic geodesic tracking it outside Margulis tubes. A relative version of the criterion of Lemma 2.8 can now be used to prove the existence of Cannon-Thurston maps.

4.3 Amalgamation Geometry. Again, as in Definition 4.4, start with a fixed closed hyperbolic surface S, a collection of simple closed curves $\mathcal{C}$ and set $I = [0,3]$. Perform Dehn surgeries on the Margulis tubes corresponding to $\mathcal{C}$ as before. Let $K = S \times [1,2] \subset S \times [0,3]$ and let $K^c = (S \times I - \cup_i N_\epsilon(\sigma_i) \times [1,2])$. Instead of fixing the product metric on the complement K^c of Margulis tubes in K, allow these complementary components to have *arbitrary geometry* subject only to the restriction that the geometries of $S \times \{1,2\}$ are fixed. Equip $S \times [0,1]$ and $S \times [2,3]$ with the product metrics. The resulting block is said to be a *block of amalgamation geometry*. After lifting to the universal cover, complements of Margulis tubes in the lifts $\widetilde{S} \times [1,2]$ are termed *amalgamation components*.

Definition 4.7. *An end E of a hyperbolic 3-manifold M has* amalgamated geometry *if*

1. *it is bi-Lipschitz homeomorphic to a model manifold E_m consisting of gluing $L-$thick and amalgamation geometry blocks end-to-end (for some L).*
2. *Amalgamation components are (uniformly) quasiconvex in $\widetilde{E}_m$.*

To construct the ladder $\mathcal{L}_\lambda$ we electrify amalgamation components as well as Margulis tubes. This ensures that in the electric metric,

1. Each amalgamation block has bounded geometry
2. The mapping class element taking $S \times \{1\}$ to $S \times \{2\}$ induces an isometry of the electrified metrics.

Quasiconvexity of $\mathcal{L}_\lambda$ in the electric metric now follows as before. To recover the data of hyperbolic geodesics from quasigeodesics lying close to $\mathcal{L}_\lambda$, we use (uniform) quasiconvexity of amalgamation components and existence of Cannon-Thurston maps follows.

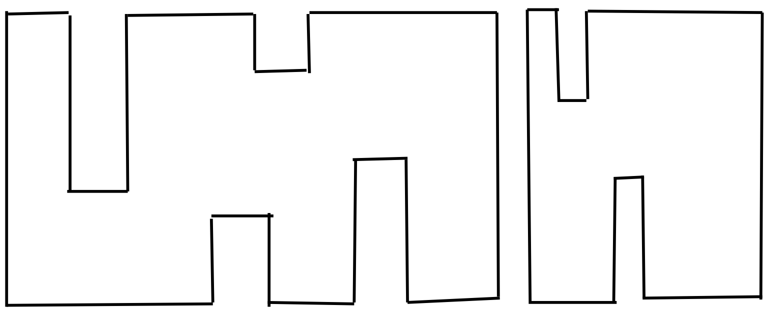
4.4 Split Geometry. We need now to relax the assumption that the boundary components of model blocks are of (uniformly) bounded geometry. Roughly speaking, split geometry is a generalization of amalgamation geometry where

1. A Margulis tube is allowed to travel through a uniformly bounded number of contiguous blocks and split them.

2. The complementary pieces, now called split components, are quasiconvex in a somewhat weaker sense (see Definition 4.8 below).

Each split component is allowed to contain Margulis tubes, called *hanging tubes* that do not go all the way across from the top to the bottom, i.e. they do not split both $\Sigma_i^s, \Sigma_{i+1}^s$.

A split component $B^s \subset B = S \times I$ is topologically a product $S^s \times I$ for some, necessarily connected $S^s (\subset S)$. However, the upper and lower boundaries of B^s need only be be split subsurfaces of S^s to allow for hanging tubes starting or ending (but not both) within the split block.



Split Block with hanging tubes

An end E of a hyperbolic 3-manifold M has *weak split geometry* if it is bi-Lipschitz homeomorphic to a model manifold E_m consisting of gluing $L-$thick and split blocks as above end-to-end (for some L). Electrifying split components as in Section 4.3, we obtain a new electric metric called the *graph metric* d_G on E.

Definition 4.8. *A model of weak split geometry is said to be of* split geometry *if the convex hull of each split component has uniformly bounded d_G-diameter.*

5 Cannon-Thurston Maps for Kleinian groups and Applications

5.1 Cannon-Thurston maps for degenerate manifolds. Let M be a hyperbolic 3-manifold homotopy equivalent to a closed hyperbolic surface S. Once we establish that M has split geometry, the proof proceeds as in Section 4.3 by electrifying split components, constructing a hyperbolic ladder $\mathcal{L}_\lambda$ and finally recovering a hyperbolic geodesic from an electric one. We shall therefore dwell in this subsection on showing that any degenerate end has split geometry. We shall do this under two simplifying assumptions, directing the reader to Mj [2014a] (especially the Introduction) for a more detailed road-map.

We borrow extensively from the hierarchy and model manifold terminology and technology of Masur and Y. N. Minsky [2000] and Y. Minsky [2010]. The model manifold of Y. Minsky [2010] and Brock, Canary, and Y. N. Minsky [2012] furnishes a resolution, or equivalently, a sequence $\{P_m\}$ of pants decompositions of S exiting E and hence a *hierarchy path*. Let τ_m denote the simple multicurve on S constituting P_m. The P_m in turn furnish split level surfaces $\{S_m\}$ exiting E: a *split level surface* S_m is a collection of (nearly) totally geodesic embeddings of the pairs of pants comprising P_m. Next, corresponding to the hierarchy path $\{\tau_m\}$, there is a tight geodesic in $\mathcal{C}(S)$ consisting of the bottom geodesic $\{\eta_i\}$ of the hierarchy. We proceed to extract a subsequence of the resolution τ_m using the bottom geodesic $\{\eta_i\}$ under two key simplifying assumptions:

1. For all i, the length of exactly one curve in η_i is sufficiently small, less than the Margulis constant in particular. Call it η_i for convenience.

2. Let S_i correspond to the first occurrence of the vertex η_i in the resolution τ_m. Assume further that the S_i's are actually split surfaces and not just split level surfaces, i.e. they all have injectivity radius uniformly bounded below,

It follows that the Margulis tube η_i splits both S_i and S_{i+1} and that the tube T_i is trapped entirely between S_i and S_{i+1}. The product region B_i between S_i and S_{i+1} is therefore a split block for all i and T_i splits it. The model manifold thus obtained is one of weak split geometry. In a sense, this is a case of 'pure split geometry', where all blocks have a split geometry structure (no thick blocks). To prove that the model is indeed of split geometry, it remains to establish the quasiconvexity condition of Definition 4.8.

Let K be a split component and $\widetilde{K}$ an elevation to $\widetilde{E}$. Let v be a boundary short curve for the split component and let T_v be the Margulis tube corresponding to v abutting K. Denote the hyperbolic convex hull by $CH(\widetilde{K})$ and pass back to a quotient in M. A crucial observation that is needed here is the fact that any pleated surface has bounded d_G-diameter. This is because thin parts of pleated surfaces lie inside Margulis tubes that get electrified in the graph metric. It therefore suffices to show that any point in $CH(K)$ lies close to a pleated surface passing near the fixed tube T_v. This last condition follows from the Brock-Bromberg drilling theorem Brock and Bromberg [2004] and the fact that the convex core of a quasi-Fuchsian group is filled by pleated surfaces Fan [1997]. This completes our sketch of a proof of the following main theorem of Mj [2014a]:
Theorem 2.5: Let $\rho : \pi_1(S) \to PSL_2(\mathbb{C})$ be a simply or doubly degenerate (closed) surface Kleinian group. Then a Cannon-Thurston map exists.

It follows that the limit set of $\rho(\pi_1(S))$ is a continuous image of S^1 and is therefore locally connected. As a first application of Theorem 2.5, we shall use the following Theorem of Anderson and Maskit [1996] to prove that connected limit sets of Kleinian groups without parabolics are locally connected.

Theorem 5.1. *Anderson and Maskit [ibid.] Let Γ be an analytically finite Kleinian group with connected limit set. Then the limit set $\Lambda(\Gamma)$ is locally connected if and only if every simply degenerate surface subgroup of Γ without accidental parabolics has locally connected limit set.*

Combining the remark after Theorem 2.5 with Theorem 5.1, we immediately have the following affirmative answer to Question 2.4.

Theorem 5.2. *Let Γ be a finitely generated Kleinian group without parabolics and with a connected limit set Λ. Then Λ is locally connected.*

Theorem 2.5 can be extended to punctured surfaces Mj [2014a] and this allows Theorem 5.2 to be generalized to arbitrary finitely generated Kleinian groups.

5.2 Finitely generated Kleinian groups. In Mj [2014b], we show that the point preimages of the Cannon-Thurston map for a simply or doubly degenerate surface Kleinian group given by Theorem 2.5 corresponds to end-points of leaves of ending laminations. In particular, the ending lamination corresponding to a degenerate end can be recovered from the Cannon-Thurston map. This was extended further in Das and Mj [2016] and Mj [2017a] to obtain the following general version for finitely generated Kleinian groups.

Theorem 5.3. *Mj [ibid.] Let G be a finitely generated Kleinian group. Let $i : \Gamma_G \to \mathbb{H}^3$ be the natural identification of a Cayley graph of G with the orbit of a point in $\mathbb{H}^3$. Further suppose that each degenerate end of $\mathbf{H}^3/G$ can be equipped with a Minsky model Y. Minsky [2010].*[2] *Then i extends continuously to a Cannon-Thurston map $\hat{i} : \widehat{\Gamma_G} \to \mathbb{D}^3$, where $\widehat{\Gamma_G}$ denotes the (relative) hyperbolic compactification of Γ_G.*

Let ∂i denote the restriction of $\hat{i}$ to the boundary $\partial \Gamma_G$ of Γ_G. Let E be a degenerate end of $N^h = \mathbb{H}^3/G$ and $\widetilde{E}$ a lift of E to $\widetilde{N^h}$ and let M_{gf} be an augmented Scott core of N^h. Then the ending lamination $\mathfrak{L}_E$ for the end E lifts to a lamination on $\widetilde{M_{gf}} \cap \widetilde{E}$. Each such lift $\mathfrak{L}$ of the ending lamination of a degenerate end defines a relation $\mathfrak{R}_{\mathfrak{L}}$ on the (Gromov) boundary $\partial \widetilde{M_{gf}}$ (or equivalently, the relative hyperbolic boundary $\partial_r \Gamma_G$ of Γ_G), given by $a \mathfrak{R}_{\mathfrak{L}} b$ iff a, b are end-points of a leaf of $\mathfrak{L}$. Let $\{\mathfrak{R}_i\}$ be the entire collection of relations on $\partial \widetilde{M_{gf}}$ obtained this way. Let $\mathfrak{R}$ be the transitive closure of the union $\bigcup_i \mathfrak{R}_i$. Then $\partial i(a) = \partial i(b)$ iff $a \mathfrak{R} b$.

[2] This hypothesis is satisfied for all ends without parabolics as well as for ends incompressible away from cusps - see Y. Minsky [2010] and Brock, Canary, and Y. N. Minsky [2012] and Mj [2017a, Appendix]. That the hypothesis is satisfied in general would follow from unpublished work of B. H. Bowditch [2016] and B. Bowditch [2005].

5.3 Primitive Stable Representations. In Y. N. Minsky [2013] Minsky introduced an open subset of the $PSL_2(\mathbb{C})$ character variety for a free group, properly containing the Schottky representations, on which the action of the outer automorphism group is properly discontinuous. He called these *primitive stable representations.* Let F_n be a free group of rank n. An element of F_n is primitive if it is an element of a free generating set. Let $\mathcal{P} = \cdots www \cdots$ be the set of bi-infinite words with w cyclically reduced primitive. A representation $\rho : F_n \to PSL_2(\mathbb{C})$ is primitive stable if all elements of $\mathcal{P}$ are mapped to uniform quasigeodesics in $\mathbb{H}^3$.

Minsky conjectured that primitive stable representations are characterized by the feature that every component of the ending lamination is blocking.

Using Theorem 5.3, Jeon and Kim [2010], and Jeon, Kim, Ohshika, and Lecuire [2014] resolved this conjecture. We briefly sketch their argument for a degenerate free Kleinian group without parabolics.

Let $\{D_1, \cdots, D_n\} = \mathfrak{D}$ be a finite set of essential disks on a handlebody H cutting H into a 3-ball. A free generating set of F_n is dual to $\mathfrak{D}$. For a lamination $\mathfrak{L}$, the *Whitehead graph* $Wh(\Lambda, \mathfrak{D})$ is defined as follows. Cut ∂H along $\partial \mathfrak{D}$ to obtain a sphere with $2n$ holes, labeled by $D_i^{\pm}$. The vertices of $Wh(\mathfrak{L}, \mathfrak{D})$ are the boundary circles of ∂H, with an edge whenever two circles are joined by an arc of $\mathfrak{L} \setminus \mathfrak{D}$. For the ending lamination $\mathfrak{L}_E$ of a degenerate free group without parabolics, $Wh(\mathfrak{L}_E, \Delta)$ is connected and has no cutpoints.

Let ρ_E be the associated representation. If ρ is not primitive stable, then there exists a sequence of primitive cyclically reduced elements w_n such that $\rho(w_n^*)$ is not an $n-$ quasigeodesic. After passing to a subsequence, w_n and hence w_n^* converges to a bi-infinite geodesic w_∞ in the Cayley graph with two distinct end points w_+, w_- in the Gromov boundary of F_n. The Cannon-Thurston map identifies w_+, w_-. Hence by Theorem 5.3 they are either the end points of a leaf of $\mathfrak{L}_E$ or ideal end-points of a complementary ideal polygon of $\mathfrak{L}_E$. It follows therefore that $Wh(w_\infty, \mathfrak{D})$ is connected and has no cutpoints. Since w_n's converge to w_∞, $Wh(w_n, \mathfrak{D})$ is connected and has no cutpoints for large enough n. A Lemma due to Whitehead says that if $Wh(w_n, \mathfrak{D})$ is connected and has no cutpoints, then w_n cannot be primitive, a contradiction.

5.4 Discreteness of Commensurators. In C. Leininger, Long, and Reid [2011] and Mj [2011], Theorems 2.5 and 5.3 are used to prove that commensurators of finitely generated, infinite covolume, Zariski dense Kleinian groups are discrete. The basic fact that goes into the proof is that commensurators preserve the structure of point pre-images of Cannon-Thurston maps. The point pre-image structure is known from Theorem 5.3.

5.5 Radial and horospherical limit sets.

Definition 5.4. *A point $\xi \in S^2$ is a radial or conical limit point of Kleinian group Γ, if for any base-point $o \in \widetilde{M}$ and any geodesic γ_ξ ending at ξ, there exist $C \geq 0$ and infinitely many translates $g.o \in N_C(\gamma_\xi)$, $g \in \Gamma$.*

A point $\xi \in S^2$ is a horospherical limit point of Kleinian group Γ, if for any base-point $o \in \widetilde{M}$ and any horoball B_ξ based at ξ, there exist infinitely many translates $g.o \in B_\xi$, $g \in \Gamma$.

The collection of radial (resp. horospherical) limit points of Γ is called the radial (resp. horospherical) limit set of Γ and is denoted by Λ_r (resp. Λ_h).

The multiple limit set Λ_m consists of those point of S^2 that have more than one pre-image under the Cannon-Thurston map.

Several authors M. Kapovich [1995], Gerasimov [2012], and Jeon, I. Kapovich, C. Leininger, and Ohshika [2016] worked on the relationship between Λ_m and Λ_r. They concluded that the conical limit set is strictly contained in the set of injective points of the Cannon-Thurston map, i.e. $\Lambda_r \subset \Lambda_m^c$, but the inclusion is proper.

In Lecuire and Mj [2016], we showed:

Theorem 5.5. $\Lambda_m^c = \Lambda_h$.

5.6 Motions of limit sets. We discuss the following question in this section, which paraphrases the second part of of Thurston [1982, Problem 14]. A detailed survey appears in Mj [2017b].

Question 5.6. *Let G_n be a sequence of Kleinian groups converging to a Kleinian group G. Does the corresponding dynamics of G_n on the Riemann sphere S^2 converge to the dynamics of G on S^2?*

To make Question 5.6 precise, we need to make sense of 'convergence' both for Kleinian groups and for their dynamics on S^2. There are three different notions of convergence for Kleinian groups.

Definition 5.7. *Let $\rho_i : H \to PSL_2(\mathbb{C})$ be a sequence of Kleinian groups. We say that that ρ_i converges to ρ_∞* algebraically *if for all $h \in H$, $\rho_i(h) \to \rho_\infty(h)$.*

Let $\rho_j : H \to PSL_2(\mathbb{C})$ be a sequence of discrete, faithful representations of a finitely generated, torsion-free, nonabelian group H. If $\{\rho_j(H)\}$ converges as a sequence of closed subsets of $PSL_2(\mathbb{C})$ to a torsion-free, nonabelian Kleinian group Γ, Γ is called the geometric limit of the sequence.

$G_i (= \rho_i(H))$ converges strongly *to $G (= \rho_\infty(H))$ if the convergence is both geometric and algebraic.*

Question 5.6 then splits into the following three questions.

Question 5.8. *1. If $G_n \to G$ geometrically, then do the corresponding limit sets converge in the Hausdorff topology on S^2?*

2. If $G_n \to G$ strongly then do the corresponding Cannon-Thurston maps converge uniformly?

3. If $G_n \to G$ algebraically then do the corresponding Cannon-Thurston maps converge pointwise?

We give the answers straight off and then proceed to elaborate.

Answers 5.9. *1. The answer to Question 5.8 (1) is Yes.*

2. The answer to Question 5.8 (2) is Yes.

3. The answer to Question 5.8 (3) is No, in general.

The most general answer to Question 5.8 (1) is due to R. A. Evans [2000], R. Evans [2004]:

Theorem 5.10. *R. A. Evans [2000], R. Evans [2004] Let $\rho_n : H \to G_n$ be a sequence of weakly type-preserving isomorphisms from a geometrically finite group H to Kleinian groups G_n with limit sets Λ_n, such that ρ_n converges algebraically to $\rho_\infty : H \to G^a_\infty$ and geometrically to G^g_∞. Let Λ_a and Λ_g denote the limit sets of G^a_∞ and G^g_∞. Then $\Lambda_n \to \Lambda_g$ in the Hausdorff metric. Further, the sequence converges strongly if and only $\Lambda_n \to \Lambda_a$ in the Hausdorff metric.*

The answer to Question 5.8 (2) is due to the author and Series Mj and Series [2017] in the case that $H = \pi_1(S)$ for a closed surface S of genus greater than one. This can be generalized to arbitrary finitely generated Kleinian groups as in Mj [2017b]:

Theorem 5.11. *Let H be a fixed group and $\rho_n(H) = \Gamma_n$ be a sequence of geometrically finite Kleinian groups converging strongly to a Kleinian group Γ. Let M_n and M_∞ be the corresponding hyperbolic manifolds. Let K be a fixed complex with fundamental group H.*

Consider embeddings $\phi_n : K \to M_n, n = 1, \cdots, \infty$ such that the maps ϕ_n are homotopic to each other by uniformly bounded homotopies (in the geometric limit). Then Cannon-Thurston maps for $\widetilde{\phi_n}$ exist and converge uniformly to the Cannon-Thurston map for $\widetilde{\phi_\infty}$.

Finally we turn to Question 5.8 (3), which turns out to be the subtlest. In Mj and Series [2013] we showed that the answer to Question 5.8 (3) is 'Yes' if the geometric limit is geometrically finite. We illustrate this with a concrete example due to Kerckhoff and Thurston Kerckhoff and Thurston [1990]

Theorem 5.12. *Fix a closed hyperbolic surface S and a simple closed geodesic σ on it. Let tw^i denote the automorphism of S given by an i-fold Dehn twist along σ. Let G_i be the quasi-Fuchsian group given by the simultaneous uniformization of $(S, tw^i(S))$. Let G_∞ denote the geometric limit of the G_i's. Let S_{i-} denote the lower boundary component of the convex core of G_i, $i = 1, \cdots, \infty$ (including ∞). Let $\phi_i : S \to S_{i-}$ be such that if $0 \in \mathbb{H}^2 = \widetilde{S}$ denotes the origin of $\mathbb{H}^2$ then $\widetilde{\phi_i}(0)$ lies in a uniformly bounded neighborhood of $0 \in \mathbb{H}^3 = \widetilde{M_i}$. We also assume (using the fact that M_∞ is a geometric limit of M_i's) that S_{i-}'s converge geometrically to $S_{\infty-}$. Then the Cannon-Thurston maps for $\widetilde{\phi_i}$ converge pointwise, but not uniformly, on $\partial\mathbb{H}^2$ to the Cannon-Thurston map for $\widetilde{\phi_\infty}$.*

However, if the geometric limit is geometrically infinite, then the answer to Question 5.8 (3) may be negative. We illustrate this with certain examples of geometric limits constructed by Brock in Brock [2001].

Theorem 5.13. *Mj and Series [2017] Fix a closed hyperbolic surface S and a separating simple closed geodesic σ on it, cutting S up into two pieces S_- and S_+. Let ϕ denote an automorphism of S such that $\phi|_{S_-}$ is the identity and $\phi|_{S_+} = \psi$ is a pseudo-Anosov of S_+ fixing the boundary. Let G_i be the quasi-Fuchsian group given by the simultaneous uniformization of $(S, \phi^i(S))$. Let G_∞ denote the geometric limit of the G_i's. Let S_{i0} denote the lower boundary component of the convex core of G_i, $i = 1, \cdots, \infty$ (including ∞). Let $\phi_i : S \to S_{i0}$ be such that if $0 \in \mathbb{H}^2 = \widetilde{S}$ denotes the origin of $\mathbb{H}^2$ then $\widetilde{\phi_i}(0)$ lies in a uniformly bounded neighborhood of $0 \in \mathbb{H}^3 = \widetilde{M_i}$. We also assume (using the fact that M_∞ is a geometric limit of M_i's) that S_{i0}'s converge geometrically to $S_{\infty 0}$.*

Let Σ be a complete hyperbolic structure on S_+ such that σ is homotopic to a cusp on Σ. Let $\mathfrak{L}$ consist of pairs (ξ_-, ξ) of ideal endpoints (on $\mathbb{S}^1_\infty$) of stable leaves λ of the stable lamination of ψ acting on $\widetilde{\Sigma}$. Also let $\partial\widetilde{\mathfrak{H}}$ denote the collection of ideal basepoints of horodisks given by lifts (contained in $\widetilde{\Sigma}$) of the cusp in Σ corresponding to σ. Let

$$\Xi = \{\xi : \text{There exists } \xi_- \text{ such that } (\xi_-, \xi) \in \mathfrak{L}; \xi_- \in \partial\widetilde{\mathfrak{H}}\}.$$

Let $\partial\phi_i$, $i = 1 \cdots, \infty$ denote the Cannon-Thurston maps for $\widetilde{\phi_i}$. Then

1. *$\partial\phi_i(\xi)$ does not converge to $\partial\phi_\infty(\xi)$ for $\xi \in \Xi$.*
2. *$\partial\phi_i(\xi)$ converges to $\partial\phi_\infty(\xi)$ for $\xi \notin \Xi$.*

In Mj and Ohshika [2017], we identify the exact criteria that lead to the discontinuity phenomenon of Theorem 5.13.

6 Gromov-Hyperbolic groups

6.1 Applications and Generalizations.

6.1.1 Normal subgroups and trees. The ladder construction of Section 3.2 has been generalized considerably. We work in the context of an exact sequence $1 \to N \to G \to Q \to 1$, with N hyperbolic and G finitely presented. We observe that for the proof of Theorem 3.5 to go through it suffices to have a qi-section of Q into G to provide a 'coarse transversal' to flow. Such a qi-section was shown to exist by Mosher [1996]. We then obtain the following.

Theorem 6.1. *Mitra [1998a] Let G be a hyperbolic group and let H be a hyperbolic normal subgroup that is normal in G. Then the inclusion of Cayley graphs $i : \Gamma_H \to \Gamma_G$ gives a Cannon-Thurston map $\hat{i} : \widehat{\Gamma_H} \to \widehat{\Gamma_G}$.*

The ladder construction can also be generalized to the general framework of a tree of hyperbolic metric spaces.

Theorem 6.2. *Mitra [1998b] Let (X, d) be a tree (T) of hyperbolic metric spaces satisfying the qi-embedded condition (i.e. edge space to vertex space inclusions are qi-embeddings). Let v be a vertex of T and (X_v, d_v) be the vertex space corresponding to v. If X is hyperbolic then the inclusion $i : X_v \to X$ gives a Cannon-Thurston map $\hat{i} : \widehat{X_v} \to \widehat{X}$.*

Theorem 6.1 was generalized by the author and Sardar to a purely coarse geometric context, where no group action is present. The relevant notion is that of a metric bundle for which we refer the reader to Mj and Sardar [2012]. Roughly speaking, the data of a metric bundle consists of vertex and edge spaces as in the case of a tree of spaces, with two notable changes:

1. The base T is replaced by an arbitrary graph B.
2. All edge-space to vertex space maps are quasi-isometries rather than just quasi-isometric embeddings.

With these modifications in place we have the following generalizations of Mosher's qi-section Lemma Mosher [1996] and Theorem 6.1:

Theorem 6.3. *Mj and Sardar [2012] Suppose $p : X \to B$ is a metric graph bundle satisfying the following:*

1. *B is a Gromov hyperbolic graph.*
2. *Each fiber F_b, for b a vertex of B is δ-hyperbolic (for some $\delta > 0$) with respect to the path metric induced from X.*
3. *The barycenter maps $\partial^3 F_b \to F_b$, $b \in B$, sending a triple of distinct points on the boundary ∂F_b to their centroid, are (uniformly, independent of b) coarsely surjective.*

4. X is hyperbolic.

Then there is a qi-section $B \to X$. The inclusion map $i_b : F_b \to X$ gives a Cannon-Thurston map $\hat{i} : \widehat{F_b} \to \widehat{X}$.

6.2 Point pre-images: Laminations. In Section 3.1, it was pointed out that the Cannon-Thurston map $\hat{i}$ identifies $p, q \in S^1$ if and only if p, q are end-points of a leaf or ideal end-points of a complementary ideal polygon of the stable or unstable lamination.

In Mitra [1997a] an algebraic theory of ending laminations was developed based on Thurston's theory Thurston [1980]. The theory was developed in the context of a normal hyperbolic subgroup of a hyperbolic group G and used to give an explicit structure for the Cannon-Thurston map in Theorem 6.1.

Definition 6.4. *Bestvina, Feighn, and Handel [1997], Coulbois, Hilion, and Lustig [2007, 2008a,b], I. Kapovich and Lustig [2010, 2015], and Mitra [1997a] An* algebraic lamination *for a hyperbolic group H is an H-invariant, flip invariant, closed subset $\mathcal{L} \subseteq \partial^2 H = (\partial H \times \partial H \setminus \Delta)/\sim$, where $(x, y) \sim (y, x)$ denotes the flip and Δ the diagonal in $\partial H \times \partial H$.*

Let
$$1 \to H \to G \to Q \to 1$$
be an exact sequence of finitely presented groups with H, G hyperbolic. It follows by work of Mosher [1996] that Q is hyperbolic. In Mitra [1997a], we construct algebraic ending laminations naturally parametrized by points in the boundary ∂Q. We describe the construction now.

Every element $g \in G$ gives an automorphism of H sending h to $g^{-1}hg$ for all $h \in H$. Let $\phi_g : \mathcal{V}(\Gamma_H) \to \mathcal{V}(\Gamma_H)$ be the resulting bijection of the vertex set $\mathcal{V}(\Gamma_H)$ of Γ_H. This induces a map Φ_g sending an edge $[a, b] \subset \Gamma_H$ to a geodesic segment joining $\phi_g(a), \phi_g(b)$.

For some (any) $z \in \partial\Gamma_Q$ we shall describe an algebraic ending lamination Λ_z. Fix such a z and let

1. $[1, z) \subset \Gamma_Q$ be a geodesic ray starting at 1 and converging to $z \in \partial\Gamma_Q$.

2. $\sigma : Q \to G$ be a qi section.

3. z_n be the vertex on $[1, z)$ such that $d_Q(1, z_n) = n$.

4. $g_n = \sigma(z_n)$.

For $h \in H$, let $\mathcal{S}_n^h$ be the H–invariant collection of all free homotopy representatives (or equivalently, shortest representatives in the same conjugacy class) of $\phi_{g_n^{-1}}(h)$ in Γ_H.

Identifying equivalent geodesics in $\mathcal{S}_n^h$ one obtains a subset $\mathbb{S}_n^h$ of unordered pairs of points in $\widehat{\Gamma_H}$. The intersection with $\partial^2 H$ of the union of all subsequential limits (in the Hausdorff topology) of $\{\mathbb{S}_n^h\}$ is denoted by Λ_z^h.

Definition 6.5. *The set of* algebraic ending laminations corresponding to $z \in \partial\Gamma_Q$ *is given by*

$$\Lambda_{EL}(z) = \bigcup_{h \in H} \Lambda_z^h.$$

Definition 6.6. *The set Λ of all algebraic ending laminations is defined by*

$$\Lambda_{EL} = \bigcup_{z \in \partial\Gamma_Q} \Lambda_{EL}(z).$$

The following was shown in Mitra [1997a]:

Theorem 6.7. *The Cannon-Thurston map $\hat{i}$ of Theorem 6.1 identifies the end-points of leaves of Λ_{EL}. Conversely, if $\hat{i}(p) = \hat{i}(q)$ for $p \neq q \in \partial\Gamma_H$, then some bi-infinite geodesic having p, q as its end-points is a leaf of Λ_{EL}.*

6.2.1 Finite-to-one. The classical Cannon-Thurston map of Theorem 2.2 is finite-to-one. Swarup asked (cf. Bestvina's Geometric Group Theory problem list Bestvina [2004, Prolem 1.20]) if the Cannon-Thurston maps of Theorem 6.1 are also finite-to-one. Kapovich and Lustig answered this in the affirmative in the following case.

Theorem 6.8. *I. Kapovich and Lustig [2015] Let $\phi \in Out(F_N)$ be a fully irreducible hyperbolic automorphism. Let $G_\phi = F_N \rtimes_\phi \mathbb{Z}$ be the associated mapping torus group. Let ∂i denote the Cannon-Thurston map of Theorem 6.1 in this case. Then for every $z \in \partial G_\phi$, the cardinality of $(\partial i)^{-1}(z)$ is at most $2N$.*

Bestvina, Feighn, and Handel [1997] define a closely related set Λ_{BFH} of algebraic laminations in the case covered by Theorem 6.8 using train-track representatives of free group automorphisms. Any algebraic lamination $\mathcal{L}$ defines a relation $\mathcal{R}_\mathcal{L}$ on ∂F_N by $a\mathcal{R}_\mathcal{L}b$ if $(a,b) \in \mathcal{L}$. The transitive closure of $\mathcal{L}$ will be called its diagonal closure. In I. Kapovich and Lustig [2015], Kapovich and Lustig further show that in the case covered by Theorem 6.8, Λ_{EL} precisely equals the diagonal closure of Λ_{BFH}.

6.3 Relative hyperbolicity. The notion of a Cannon-Thurston map can be extended to the context of relative hyperbolicity. This was done in Mj and Pal [2011]. Let X and Y be relatively hyperbolic spaces, hyperbolic relative to the collections $\mathcal{H}_X$ and $\mathcal{H}_Y$ of 'horosphere-like sets' respectively. Let us denote the *horoballifications* of X and Y with respect to $\mathcal{H}_X$ and $\mathcal{H}_Y$ by $\mathcal{G}(X, \mathcal{H}_X), \mathcal{G}(Y, \mathcal{H}_Y)$ respectively (see B. H. Bowditch [2012] for details). The horoballification of an H in $\mathcal{H}_X$ or $\mathcal{H}_Y$ is denoted as H^h. Note that $\mathcal{G}(X, \mathcal{H}_X), \mathcal{G}(Y, \mathcal{H}_Y)$ are hyperbolic. The electrifications will be denoted as $\mathcal{E}(X, \mathcal{H}_X), \mathcal{E}(Y, \mathcal{H}_Y)$.

Definition 6.9. *A map $i : Y \to X$ is strictly type-preserving if*

1. *for all $H_Y \in \mathcal{H}_Y$ there exists $H_X \in \mathcal{H}_X$ such that $i(H_Y) \subset H_X$, and*

2. *images of distinct horospheres-like sets in Y lie in distinct horosphere-like sets in X.*

Let $i : Y \to X$ be a strictly type-preserving proper embedding. Then i induces a proper embedding $i_h : \mathcal{G}(Y, \mathcal{H}_Y) \to \mathcal{G}(X, \mathcal{H}_X)$.

Definition 6.10. *A Cannon-Thurston map exists for a strictly type-preserving inclusion $i : Y \to X$ of relatively hyperbolic spaces if a Cannon-Thurston map exists for the induced map $i_h : \mathcal{G}(Y, \mathcal{H}_Y) \to \mathcal{G}(X, \mathcal{H}_X)$.*

Lemma 2.8 generalizes to the following.

Lemma 6.11. *A Cannon-Thurston map for $i : Y \to X$ exists if and only if there exists a non-negative proper function $M : \mathbb{N} \to \mathbb{N}$ such that the following holds:*
Fix a base-point $y_0 \in Y$. Let $\hat{\lambda}$ in $\mathcal{E}(Y, \mathcal{H}_Y)$ be an electric geodesic segment starting and ending outside horospheres. If $\lambda^b = \hat{\lambda} \setminus \bigcup_{K \in \mathcal{H}_Y} K$ lies outside $B_N(y_0) \subset Y$, then for any electric quasigeodesic $\hat{\beta}$ joining the end points of $\hat{i}(\hat{\lambda})$ in $\mathcal{E}(X, \mathcal{H}_X)$, $\beta^b = \hat{\beta} \setminus \bigcup_{H \in \mathcal{H}_X} H$ lies outside $B_{M(N)}(i(y_0)) \subset X$.

Theorem 6.2 then generalizes to:

Theorem 6.12. *Mj and Pal [2011] Let $P : X \to T$ be a tree of relatively hyperbolic spaces satisfying the qi-embedded condition. Assume that*

1. *the inclusion maps of edge-spaces into vertex spaces are strictly type-preserving*

2. *the induced tree of electrified (coned-off) spaces continues to satisfy the qi-embedded condition*

3. *X is strongly hyperbolic relative to the family $\mathcal{C}$ of maximal cone-subtrees of horosphere-like sets.*

Then a Cannon-Thurston map exists for the inclusion of relatively hyperbolic spaces $i : X_v \to X$, where (X_v, d_{X_v}) is the relatively hyperbolic vertex space corresponding to v.

6.4 Problems. The above survey is conditioned and limited by the author's bias on the one hand and space considerations on the other. In particular we have omitted the important work on quasigeodesic foliations by Calegari [2006, 2000], Fenley [2012], and Frankel [2015] and the topology of ending lamination spaces by Gabai and others Gabai [2014, 2009], C. J. Leininger, Mj, and Schleimer [2011], and Hensel and Przytycki [2011] as this would be beyond the scope of the present article. A more detailed survey appears in Papadopoulos [2007]. We end with some open problems.

6.4.1 Higher dimensional Kleinian groups. As a test-case we propose:

Question 6.13. *Let S be a closed surface of genus greater than one and let $\Gamma = \pi_1(S)$ act freely, properly discontinuously by isometries on $\mathbf{H}^n$, $n > 3$ (or more generally a rank one symmetric space). Does a Cannon-Thurston map exist in general?*

Can the small cancellation group of Baker-Riley in Baker and T. R. Riley [2013] act geometrically on a rank one symmetric space thus giving a negative answer to Question 6.13 with surface group replaced by free group? Work of Wise [2009] and Wise [2004] guarantees linearity of such small cancellation groups.

The critical problem in trying to answer Question 6.13 is the absence of new examples in higher dimensions. It would be good to find new examples or prove that they do not exist. A version of a question due to M. Kapovich [2008] makes this more precise and indicates our current state of knowledge/ignorance:

Question 6.14. *Let S be a closed surface of genus greater than one and let Γ be a discrete subgroup of $SO(n,1)$ (or more generally a rank one Lie group)) abstractly isomorphic to $\pi_1(S)$) acting by isometries on $\mathbf{H}^n$, $n > 3$ (more generally the associated symmetric space) such that*

1. *orbits are not quasiconvex,*
2. *no element of Γ is a parabolic.*

Does ρ factor through a representation to a simply or doubly degenerate (3-dimensional) Kleinian group followed by a deformation of $SO(3,1)$ in $SO(n,1)$?

A closely related folklore question asks:

Question 6.15. *Can a closed higher dimensional $n > 3$ rank one manifold fiber? In particular over the circle?*

It is known, from the Chern-Gauss-Bonnet theorem that a $2n$ dimensional rank one manifold cannot fiber over the circle. Unpublished work of M. Kapovich [1998] shows that a complex hyperbolic 4-manifold cannot fiber over a 2-manifold.

6.4.2 Surface groups in higher rank. A topic of considerable current interest is higher dimensional Teichmüller theory and Anosov representations of surface groups Labourie [2006], M. Kapovich, Leeb, and Porti [2017], and Guéritaud, Guichard, Kassel, and Wienhard [2017]. Kapovich, Leeb and Porti give an equivalent definition of Anosov representations in purely coarse geometric terms as representations that are *asymptotic embeddings*. It will take us too far afield to define these notions precisely here. What we will say however is that if $\rho : \pi_1(S) \to \mathcal{G}$ is a discrete faithful representation into a semi-simple Lie group $\mathcal{G}$ and $\mathcal{G}/\mathcal{P} = \mathcal{B}$ is the Furstenberg boundary, then the Anosov property implies that an orbit $\rho(\pi_1(S)).o$ 'extends' to a $\rho(\pi_1(S))$−equivariant embedding of $\partial\Gamma_{\pi_1(S)}(= S^1)$ into $\mathcal{B}$. Thus the boundary map $\Delta : S^1 \to \mathcal{B}$ maybe thought of as a *higher rank Cannon-Thurston map*. For the representation to be Anosov, Δ is thus an embedding M. Kapovich, Leeb, and Porti [2017].

Question 6.16. *What class of representations do we get if we require only that Δ is continuous?*

Question 6.16 is basically asking for a 'nice' characterization of representations that admit a higher rank Cannon-Thurston map. The core problem in addressing it again boils down to finding some rich class of examples. Question 6.14 has a natural generalization to this context where we replace $SO(n, 1)$ by $\mathcal{G}$.

6.4.3 Geometric group theory. As we have seen in Section 6.1, normal hyperbolic subgroups and trees of spaces provide examples where there is a positive answer to Question 2.7. Some sporadic new examples have also been found, e.g. hydra groups Baker and T. Riley [2012]. However no systematic theory exists. In the light of the counterexample in Baker and T. R. Riley [2013], the general answer to Question 2.7 is negative. Are there necessary and/or sufficient conditions beyond Lemma 2.8 to guarantee existence of Cannon-Thurston maps?

As illustrated in Mitra [1998b] and Baker and T. Riley [2012], distortion of subgroups Gromov [1993] is irrelevant. Distortion captures the relationship between $d_H(1, h)$ with $d_G(1, h)$. On the other hand Cannon-Thurston maps capture the corresponding relationship between $d_H(1, [h_1, h_2]_H)$ with $d_G(1, [h_1, h_2]_G)$, i.e. existence of Cannon-Thurston maps is equivalent to a proper embedding of 'pairs of points' (coding geodesic segments). The function associated with such a proper embedding is closely related to the modulus of continuity of the Cannon-Thurston map Baker and T. Riley [2012].

Acknowledgments. I would like to thank Benson Farb and Thomas Koberda for several helpful comments on an earlier draft.

References

William Abikoff (1976). "Two theorems on totally degenerate Kleinian groups". *Amer. J. Math.* 98.1, pp. 109–118. MR: 0396937 (cit. on pp. 910, 911).

R. C. Alperin, Warren Dicks, and J. Porti (1999). "The boundary of the Gieseking tree in hyperbolic three-space". *Topology Appl.* 93.3, pp. 219–259. MR: 1688476 (cit. on p. 911).

James W. Anderson and Bernard Maskit (1996). "On the local connectivity of limit set of Kleinian groups". *Complex Variables Theory Appl.* 31.2, pp. 177–183. MR: 1423249 (cit. on pp. 920, 921).

O. Baker and T. R. Riley (2013). "Cannon-Thurston maps do not always exist". *Forum Math. Sigma* 1, e3, 11. MR: 3143716 (cit. on pp. 912, 930, 931).

Owen Baker and Timothy Riley (Sept. 2012). "Cannon-Thurston maps, subgroup distortion, and hyperbolic hydra". arXiv: 1209.0815 (cit. on p. 931).

M. Bestvina and M. Feighn (1992). "A combination theorem for negatively curved groups". *J. Differential Geom.* 35.1, pp. 85–101. MR: 1152226 (cit. on p. 913).

M. Bestvina, M. Feighn, and M. Handel (1997). "Laminations, trees, and irreducible automorphisms of free groups". *Geom. Funct. Anal.* 7.2, pp. 215–244. MR: 1445386 (cit. on pp. 927, 928).

Mladen Bestvina (2004). "Geometric group theory problem list" (cit. on pp. 911, 928).

Francis Bonahon (1986). "Bouts des variétés hyperboliques de dimension 3". *Ann. of Math. (2)* 124.1, pp. 71–158. MR: 847953 (cit. on p. 908).

B. H. Bowditch (2012). "Relatively hyperbolic groups". *Internat. J. Algebra Comput.* 22.3, pp. 1250016, 66. MR: 2922380 (cit. on p. 928).

– (2013). "Stacks of hyperbolic spaces and ends of 3-manifolds". In: *Geometry and topology down under*. Vol. 597. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 65–138. MR: 3186670 (cit. on p. 911).

BH Bowditch (2005). "End invariants of hyperbolic 3-manifolds". preprint, Southampton (cit. on p. 921).

Brian H Bowditch (2016). "Ending lamination theorem". Preprint, Warwick (cit. on p. 921).

Brian H. Bowditch (2007). "The Cannon-Thurston map for punctured-surface groups". *Math. Z.* 255.1, pp. 35–76. MR: 2262721 (cit. on p. 911).

Jeffrey F. Brock (2001). "Iteration of mapping classes and limits of hyperbolic 3-manifolds". *Invent. Math.* 143.3, pp. 523–570. MR: 1817644 (cit. on p. 925).

Jeffrey F. Brock and Kenneth W. Bromberg (2004). "On the density of geometrically finite Kleinian groups". *Acta Math.* 192.1, pp. 33–93. MR: 2079598 (cit. on p. 920).

Jeffrey F. Brock, Richard D. Canary, and Yair N. Minsky (2012). "The classification of Kleinian surface groups, II: The ending lamination conjecture". *Ann. of Math. (2)* 176.1, pp. 1–149. MR: 2925381 (cit. on pp. 904, 909, 915, 920, 921).

Danny Calegari (2000). "The geometry of **R**-covered foliations". *Geom. Topol.* 4, pp. 457–515. MR: 1800151 (cit. on p. 929).

– (2006). "Universal circles for quasigeodesic flows". *Geom. Topol.* 10, pp. 2271–2298. MR: 2284058 (cit. on p. 929).

James W. Cannon and William P. Thurston (1985). "Group invariant Peano curves". Preprint, Princeton (cit. on pp. 910, 911).

– (2007). "Group invariant Peano curves". *Geom. Topol.* 11, pp. 1315–1355. MR: 2326947 (cit. on pp. 910, 911).

Thierry Coulbois, Arnaud Hilion, and Martin Lustig (2007). "Non-unique ergodicity, observers' topology and the dual algebraic lamination for ℝ-trees". *Illinois J. Math.* 51.3, pp. 897–911. MR: 2379729 (cit. on p. 927).

– (2008a). "ℝ-trees and laminations for free groups. I. Algebraic laminations". *J. Lond. Math. Soc. (2)* 78.3, pp. 723–736. MR: 2456901 (cit. on p. 927).

– (2008b). "ℝ-trees and laminations for free groups. II. The dual lamination of an ℝ-tree". *J. Lond. Math. Soc. (2)* 78.3, pp. 737–754. MR: 2456902 (cit. on p. 927).

Shubhabrata Das and Mahan Mj (2016). "Semiconjugacies between relatively hyperbolic boundaries". *Groups Geom. Dyn.* 10.2, pp. 733–752. MR: 3513115 (cit. on p. 921).

Spencer Dowdall, Ilya Kapovich, and Samuel J. Taylor (2016). "Cannon-Thurston maps for hyperbolic free group extensions". *Israel J. Math.* 216.2, pp. 753–797. MR: 3557464 (cit. on p. 915).

Spencer Dowdall, Richard P. Kent IV, and Christopher J. Leininger (2014). "Pseudo-Anosov subgroups of fibered 3-manifold groups". *Groups Geom. Dyn.* 8.4, pp. 1247–1282. MR: 3314946 (cit. on p. 915).

Spencer Dowdall and Samuel J. Taylor (2017). "The co-surface graph and the geometry of hyperbolic free group extensions". *J. Topol.* 10.2, pp. 447–482. MR: 3653318 (cit. on p. 915).

– (2018). "Hyperbolic extensions of free groups". *Geom. Topol.* 22.1, pp. 517–570. arXiv: 1406.2567. MR: 3720349 (cit. on p. 915).

Richard Evans (2004). "Weakly type-preserving sequences and strong convergence". *Geom. Dedicata* 108, pp. 71–92. MR: 2112665 (cit. on p. 924).

Richard Allen Evans (2000). *Deformation spaces of hyperbolic 3-manifolds: Strong convergence and tameness*. Thesis (Ph.D.)–University of Michigan. ProQuest LLC, Ann Arbor, MI, p. 134. MR: 2700410 (cit. on p. 924).

Carol E. Fan (1997). *Injectivity radius bounds in hyperbolic convex cores*. Thesis (Ph.D.)–University of Michigan. ProQuest LLC, Ann Arbor, MI, p. 114. MR: 2695849 (cit. on p. 920).

B. Farb (1998). "Relatively hyperbolic groups". *Geom. Funct. Anal.* 8.5, pp. 810–840. MR: 1650094 (cit. on pp. 915, 917).

A. Fathi, M. Laudenbach, and V. Poenaru (1979). *Travaux de Thurston sur les surfaces*. Vol. 66. Astérisque. Séminaire Orsay, With an English summary. Société Mathématique de France, Paris, p. 284. MR: 568308 (cit. on p. 907).

Sérgio Fenley (2012). "Ideal boundaries of pseudo-Anosov flows and uniform convergence groups with connections and applications to large scale geometry". *Geom. Topol.* 16.1, pp. 1–110. MR: 2872578 (cit. on p. 929).

William J. Floyd (1980). "Group completions and limit sets of Kleinian groups". *Invent. Math.* 57.3, pp. 205–218. MR: 568933 (cit. on p. 911).

Steven Frankel (2015). "Quasigeodesic flows and sphere-filling curves". *Geom. Topol.* 19.3, pp. 1249–1262. MR: 3352235 (cit. on p. 929).

David Gabai (2009). "Almost filling laminations and the connectivity of ending lamination space". *Geom. Topol.* 13.2, pp. 1017–1041. MR: 2470969 (cit. on p. 929).

– (2014). "On the topology of ending lamination space". *Geom. Topol.* 18.5, pp. 2683–2745. MR: 3285223 (cit. on p. 929).

Victor Gerasimov (2012). "Floyd maps for relatively hyperbolic groups". *Geom. Funct. Anal.* 22.5, pp. 1361–1399. MR: 2989436 (cit. on p. 923).

M. Gromov (1987). "Hyperbolic groups". In: *Essays in group theory*. Vol. 8. Math. Sci. Res. Inst. Publ. Springer, New York, pp. 75–263. MR: 919829 (cit. on p. 911).

– (1993). "Asymptotic invariants of infinite groups". In: *Geometric group theory, Vol. 2 (Sussex, 1991)*. Vol. 182. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 1–295. MR: 1253544 (cit. on p. 931).

François Guéritaud, Olivier Guichard, Fanny Kassel, and Anna Wienhard (2017). "Anosov representations and proper actions". *Geom. Topol.* 21.1, pp. 485–584. MR: 3608719 (cit. on p. 930).

Sebastian Hensel and Piotr Przytycki (2011). "The ending lamination space of the five-punctured sphere is the Nöbeling curve". *J. Lond. Math. Soc. (2)* 84.1, pp. 103–119. MR: 2819692 (cit. on p. 929).

Woojin Jeon, Ilya Kapovich, Christopher Leininger, and Ken'ichi Ohshika (2016). "Conical limit points and the Cannon-Thurston map". *Conform. Geom. Dyn.* 20, pp. 58–80. MR: 3488025 (cit. on p. 923).

Woojin Jeon and Inkang Kim (2010). "Primitive stable representations of geometrically infinite handlebody hyperbolic 3-manifolds". *C. R. Math. Acad. Sci. Paris* 348.15-16, pp. 907–910. MR: 2677989 (cit. on p. 922).

Woojin Jeon, Inkang Kim, Ken'ichi Ohshika, and Cyril Lecuire (2014). "Primitive stable representations of free Kleinian groups". *Israel J. Math.* 199.2, pp. 841–866. MR: 3219560 (cit. on p. 922).

Yuichi Kabaya (2016). "Deformation of quasi-Fuchsian groups" (cit. on p. 906).

Ilya Kapovich and Martin Lustig (2010). "Intersection form, laminations and currents on free groups". *Geom. Funct. Anal.* 19.5, pp. 1426–1467. MR: 2585579 (cit. on p. 927).

– (2015). "Cannon-Thurston fibers for iwip automorphisms of F_N". *J. Lond. Math. Soc. (2)* 91.1, pp. 203–224. MR: 3335244 (cit. on pp. 927, 928).

M Kapovich (1995). "On absence of Ahlfors' finiteness theorem for Kleinian groups in dimension 3". In: *Algebra and analysis, Irkutsk, 1989*. Ed. by L. A. Bokut, M. Hazewinkel, and Y. G. Reshetnyak. Vol. 2, p. 163 (cit. on p. 923).

Michael Kapovich (1998). "On normal subgroups in the fundamental groups of complex surfaces". arXiv: math/9808085 (cit. on p. 930).

– (2001). *Hyperbolic manifolds and discrete groups*. Vol. 183. Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, pp. xxvi+467. MR: 1792613 (cit. on p. 907).

– (2008). "Kleinian groups in higher dimensions". In: *Geometry and dynamics of groups and spaces*. Vol. 265. Progr. Math. Birkhäuser, Basel, pp. 487–564. MR: 2402415 (cit. on p. 930).

Michael Kapovich, Bernhard Leeb, and Joan Porti (2017). "Anosov subgroups: dynamical and geometric characterizations". *Eur. J. Math.* 3.4, pp. 808–898. arXiv: 1703.01647. MR: 3736790 (cit. on pp. 930, 931).

Steven P. Kerckhoff and William P. Thurston (1990). "Noncontinuity of the action of the modular group at Bers' boundary of Teichmüller space". *Invent. Math.* 100.1, pp. 25–47. MR: 1037141 (cit. on p. 924).

Erica Klarreich (1999). "Semiconjugacies between Kleinian group actions on the Riemann sphere". *Amer. J. Math.* 121.5, pp. 1031–1078. MR: 1713300 (cit. on p. 911).

François Labourie (2006). "Anosov flows, surface groups and curves in projective space". *Invent. Math.* 165.1, pp. 51–114. MR: 2221137 (cit. on p. 930).

Cyril Lecuire and Mahan Mj (June 2016). "Horospheres in degenerate 3-manifolds". To appear in Int. Math. Res. Not. arXiv: 1606.07185 (cit. on pp. 916, 923).

Christopher J. Leininger, Mahan Mj, and Saul Schleimer (2011). "The universal Cannon-Thurston map and the boundary of the curve complex". *Comment. Math. Helv.* 86.4, pp. 769–816. MR: 2851869 (cit. on p. 929).

Christopher Leininger, Darren D. Long, and Alan W. Reid (2011). "Commensurators of finitely generated nonfree Kleinian groups". *Algebr. Geom. Topol.* 11.1, pp. 605–624. MR: 2783240 (cit. on p. 922).

H. A. Masur and Y. N. Minsky (2000). "Geometry of the complex of curves. II. Hierarchical structure". *Geom. Funct. Anal.* 10.4, pp. 902–974. MR: 1791145 (cit. on p. 919).

Yoshifumi Matsuda and Shin-ichi Oguni (2014). "On Cannon-Thurston maps for relatively hyperbolic groups". *J. Group Theory* 17.1, pp. 41–47. MR: 3176651 (cit. on p. 912).

Curtis T. McMullen (2001). "Local connectivity, Kleinian groups and geodesics on the blowup of the torus". *Invent. Math.* 146.1, pp. 35–91. MR: 1859018 (cit. on p. 911).

Yair Minsky (2010). "The classification of Kleinian surface groups. I. Models and bounds". *Ann. of Math. (2)* 171.1, pp. 1–107. MR: 2630036 (cit. on pp. 909, 915, 919–921).

Yair N. Minsky (1993). "Teichmüller geodesics and ends of hyperbolic 3-manifolds". *Topology* 32.3, pp. 625–647. MR: 1231968 (cit. on p. 916).

– (1994). "On rigidity, limit sets, and end invariants of hyperbolic 3-manifolds". *J. Amer. Math. Soc.* 7.3, pp. 539–588. MR: 1257060 (cit. on pp. 911, 915, 916).

– (2001). "Bounded geometry for Kleinian groups". *Invent. Math.* 146.1, pp. 143–192. MR: 1859020 (cit. on pp. 915, 916).

Yair N. Minsky (2013). "On dynamics of $Out(F_n)$ on $\mathrm{PSL}_2(\mathbb{C})$ characters". *Israel J. Math.* 193.1, pp. 47–70. MR: 3038545 (cit. on p. 922).

M. Mitra (1997a). "Ending laminations for hyperbolic group extensions". *Geom. Funct. Anal.* 7.2, pp. 379–402. MR: 1445392 (cit. on pp. 927, 928).

Mahan Mitra (1997b). *Maps on boundaries of hyperbolic metric spaces*. Thesis (Ph.D.)–University of California, Berkeley. ProQuest LLC, Ann Arbor, MI, p. 66. MR: 2696278 (cit. on p. 911).

– (1998a). "Cannon-Thurston maps for hyperbolic group extensions". *Topology* 37.3, pp. 527–538. MR: 1604882 (cit. on pp. 912–914, 926).

– (1998b). "Cannon-Thurston maps for trees of hyperbolic metric spaces". *J. Differential Geom.* 48.1, pp. 135–164. MR: 1622603 (cit. on pp. 911–914, 926, 931).

– (1998c). "Coarse extrinsic geometry: a survey". In: *The Epstein birthday schrift*. Vol. 1. Geom. Topol. Monogr. Geom. Topol. Publ., Coventry, pp. 341–364. MR: 1668308 (cit. on p. 911).

H Miyachi (2002). "Semiconjugacies between actions of topologically tame Kleinian groups". Preprint (cit. on p. 911).

Mahan Mj (2010). "Cannon-Thurston maps and bounded geometry". In: *Teichmüller theory and moduli problem*. Vol. 10. Ramanujan Math. Soc. Lect. Notes Ser. Ramanujan Math. Soc., Mysore, pp. 489–511. MR: 2667569 (cit. on pp. 911, 914, 915).

– (2011). "On discreteness of commensurators". *Geom. Topol.* 15.1, pp. 331–350. MR: 2776846 (cit. on p. 922).

– (2014a). "Cannon-Thurston maps for surface groups". *Ann. of Math. (2)* 179.1, pp. 1–80. MR: 3126566 (cit. on pp. 911, 915, 919–921).

– (2014b). "Ending laminations and Cannon-Thurston maps". *Geom. Funct. Anal.* 24.1. With an appendix by Shubhabrata Das and Mj, pp. 297–321. MR: 3177384 (cit. on pp. 911, 921).

– (2016). "Cannon-Thurston maps for surface groups: an exposition of amalgamation geometry and split geometry". In: *Geometry, topology, and dynamics in negative curvature*. Vol. 425. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 221–271. MR: 3497262 (cit. on pp. 911, 915).

– (2017a). "Cannon-Thurston maps for Kleinian groups". *Forum Math. Pi* 5, e1, 49. MR: 3652816 (cit. on pp. 911, 921).

– (2017b). "Motions of limit sets: A survey". In: *Proceedings of Workshop on Grothendieck–Teichmüller theories, Chern Institute, Tianjin, July 2016* (cit. on pp. 923, 924).

– (2009-2010). "Cannon-Thurston maps, i-bounded geometry and a theorem of McMullen". In: *Actes du Séminaire de Théorie Spectrale et Géometrie. Volume 28. Année 2009–2010*. Vol. 28. Sémin. Théor. Spectr. Géom. Univ. Grenoble I, Saint-Martin-d'Hères, pp. 63–107. MR: 2848212 (cit. on pp. 911, 915–917).

Mahan Mj and Ken'ichi Ohshika (Apr. 2017). "Discontinuous motions of limit sets". arXiv: 1704.00269 (cit. on p. 925).
Mahan Mj and Abhijit Pal (2011). "Relative hyperbolicity, trees of spaces and Cannon-Thurston maps". *Geom. Dedicata* 151, pp. 59–78. MR: 2780738 (cit. on pp. 928, 929).
Mahan Mj and Kasra Rafi (June 2015). "Algebraic Ending Laminations and Quasiconvexity". arXiv: 1506.08036 (cit. on p. 915).
Mahan Mj and Pranab Sardar (2012). "A combination theorem for metric bundles". *Geom. Funct. Anal.* 22.6, pp. 1636–1707. MR: 3000500 (cit. on p. 926).
Mahan Mj and Caroline Series (2013). "Limits of limit sets I". *Geom. Dedicata* 167, pp. 35–67. MR: 3128770 (cit. on p. 924).
– (2017). "Limits of limit sets II: Geometrically infinite groups". *Geom. Topol.* 21.2, pp. 647–692. MR: 3626590 (cit. on pp. 924, 925).
Lee Mosher (1996). "Hyperbolic extensions of groups". *J. Pure Appl. Algebra* 110.3, pp. 305–314. MR: 1393118 (cit. on pp. 926, 927).
Jean-Pierre Otal (2001). *The hyperbolization theorem for fibered 3-manifolds*. Vol. 7. SMF/AMS Texts and Monographs. Translated from the 1996 French original by Leslie D. Kay. American Mathematical Society, Providence, RI; Société Mathématique de France, Paris, pp. xiv+126. MR: 1855976 (cit. on p. 907).
Athanase Papadopoulos, ed. (2007). *Handbook of Teichmüller theory. Vol. I*. Vol. 11. IRMA Lectures in Mathematics and Theoretical Physics. European Mathematical Society (EMS), Zürich, pp. viii+794. MR: 2284826 (cit. on p. 929).
G. P. Scott and G. A. Swarup (1990). "Geometric finiteness of certain Kleinian groups". *Proc. Amer. Math. Soc.* 109.3, pp. 765–768. MR: 1013981 (cit. on p. 915).
Peter Scott (1978). "Subgroups of surface groups are almost geometric". *J. London Math. Soc. (2)* 17.3, pp. 555–565. MR: 0494062 (cit. on p. 915).
William P. Thurston (1980). "The geometry and topology of 3-manifolds". Lecture note (cit. on pp. 906–908, 927).
– (1982). "Three-dimensional manifolds, Kleinian groups and hyperbolic geometry". *Bull. Amer. Math. Soc. (N.S.)* 6.3, pp. 357–381. MR: 648524 (cit. on pp. 904, 909, 910, 923).
– (Jan. 1986). "Hyperbolic Structures on 3-manifolds, II: Surface groups and 3-manifolds which fiber over the circle". arXiv: math/9801045 (cit. on p. 907).
D. T. Wise (2004). "Cubulating small cancellation groups". *Geom. Funct. Anal.* 14.1, pp. 150–214. MR: 2053602 (cit. on p. 930).
Daniel T. Wise (2009). "Research announcement: the structure of groups with a quasiconvex hierarchy". *Electron. Res. Announc. Math. Sci.* 16, pp. 44–55. MR: 2558631 (cit. on p. 930).

Received 2017-09-03.

Mahan Mj
School of Mathematics
Tata Institute of Fundamental Research
1, Homi Bhabha Road
Mumbai-400005
India
mahan@math.tifr.res.in

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (937–958)

GROUPS ACTING ACYLINDRICALLY ON HYPERBOLIC SPACES

Denis V. Osin

Abstract

The goal of this article is to survey some recent developments in the study of groups acting on hyperbolic spaces. We focus on the class of *acylindrically hyperbolic groups*; it is broad enough to include many examples of interest, yet a significant part of the theory of hyperbolic and relatively hyperbolic groups can be generalized in this context. In particular, we discuss group theoretic Dehn filling and small cancellation theory in acylindrically hyperbolic groups. Many results discussed here rely on the new generalization of relative hyperbolicity based on the notion of a *hyperbolically embedded subgroup*.

1 Introduction

Suppose that a group G acts by isometries on a metric space S. If the action is sufficiently "nice", many properties of G can be revealed by studying the geometric structure of G-orbits in S. This approach works especially well if S satisfies certain negative curvature condition.

Systematic research in this direction began in late 1980s when Gromov [1987] introduced the notion of an abstract hyperbolic metric space. Groups acting properly and cocompactly on hyperbolic spaces are called *word hyperbolic*. More generally, replacing properness with its relative analogue modulo a fixed collection of subgroups leads to the notion of a *relatively hyperbolic group*. The study of hyperbolic and relatively hyperbolic groups was initiated by Gromov [ibid.] and since then it has been one of the most active areas of research in geometric group theory.

A further generalization, the class of *acylindrically hyperbolic groups*, was suggested by Osin [2016] and received considerable attention in the past few years. It includes many examples of interest: non-elementary hyperbolic and relatively hyperbolic groups, all but

The author was supported by the NSF grant DMS-1612473.
MSC2010: primary 20F65; secondary 20F67, 20F69.

finitely many mapping class groups of punctured closed surfaces, $Out(F_n)$ for $n \geq 2$, most 3-manifold groups, groups of deficiency at least 2, and the Cremona group of birational transformations of the complex projective plane, just to name a few. On the other hand, the property of being acylindrically hyperbolic is strong enough to allow one to apply powerful geometric techniques.

A significant part of the theory of relatively hyperbolic groups can be generalized to acylindrically hyperbolic groups using the notion of a *hyperbolically embedded collection of subgroups* introduced by Dahmani, Guirardel, and Osin [2017]. In particular, this notion provides a suitable framework for developing a group theoretic version of Thurston's theory of hyperbolic Dehn filling in 3-manifolds. Group theoretic Dehn filling was originally studied in the context of relatively hyperbolic groups by Groves and Manning [2008] and Osin [2007]. Recently it was used to obtain several deep results (most notably, it was employed in the proof of the virtual Haken conjecture by Agol [2013]). Yet another powerful tool is *small cancellation theory*, which can be used to prove various embedding theorems and to construct groups with unusual properties, see Hull [2016] and Osin [2010].

The main purpose of this paper is to survey the recent progress in the study of acylindrically hyperbolic groups and their hyperbolically embedded subgroups. In the next section we briefly discuss equivalent definitions, main examples, and basic properties of acylindrically hyperbolic groups. Hyperbolically embedded subgroups are discussed in Section 3. Section 4 is devoted to group theoretic Dehn filling. An informal discussion of small cancellation theory and a survey of some application is given in Section 5.

2 Acylindrically hyperbolic groups

2.1. Hyperbolic spaces and group actions. We begin by recalling basic definitions and general results about groups acting on hyperbolic spaces. Our main reference is Gromov [1987]; additional details can be found in Bridson and Haefliger [1999] and Ghys and de la Harpe [1990]. All group actions on metric spaces discussed in this paper are assumed to be isometric by default.

Definition 2.1. A metric space S is *hyperbolic* if it is geodesic and there exists $\delta \geq 0$ such that for any geodesic triangle Δ in S, every side of Δ is contained in the union of the δ-neighborhoods of the other two sides.

Example 2.2. Every bounded space S is hyperbolic with $\delta = \operatorname{diam}(S)$. Every tree is hyperbolic with $\delta = 0$. $\mathbb{H}^n$ is hyperbolic for every $n \in \mathbb{N}$. On the other hand, $\mathbb{R}^n$ is not hyperbolic for $n \geq 2$.

Given a hyperbolic space S, we denote by ∂S its *Gromov boundary*. We do not assume that the space is proper and therefore the boundary is defined as the set of equivalence classes of sequences of points convergent at infinity; for details we refer to Gromov [1987, Section 1.8]. The union $\widehat{S} = S \cup \partial S$ is a completely metrizable Hausdorff topological space containing S as a dense subset.

Example 2.3. The Gromov boundary of a bounded space is empty. $\partial(\mathbb{H}^n) = \mathbb{S}^{n-1}$. The boundary of an n-regular tree is the Cantor set if $n \geq 3$ and consists of two points if $n = 2$.

Let G be a group acting (isometrically) on a hyperbolic space S. This action extends to an action on $\widehat{S}$ by homeomorphisms. We denote by $\Lambda(G)$ the *limit set* of G, that is, the set of accumulation points of a G-orbit on ∂S. Thus

$$\Lambda(G) = \overline{Gs} \cap \partial S,$$

where $s \in S$ and $\overline{Gs}$ is the closure of the corresponding orbit. In fact, this definition is independent of the choice of $s \in S$. Given an element $g \in G$, we denote $\Lambda(\langle g \rangle)$ simply by $\Lambda(g)$ and call it the *limit set* of g.

Similarly to the classification of elements of $PSL(2, \mathbb{R}) = \operatorname{Isom} \mathbb{H}^2$, we have the following classification of isometries of abstract hyperbolic spaces.

Definition 2.4. An element $g \in G$ is called *elliptic* if $\Lambda(g) = \varnothing$ (equivalently, all orbits of $\langle g \rangle$ are bounded), *parabolic* if $|\Lambda(g)| = 1$; and *loxodromic* if $|\Lambda(g)| = 2$. Equivalently, an element $g \in G$ is loxodromic if the map $\mathbb{Z} \to S$ defined by $n \mapsto g^n s$ is a quasi-isometric embedding for every $s \in S$; in turn, this is equivalent to the existence of $c > 0$ such that $\mathrm{d}_S(s, g^n s) \geq c|n|$ for all $n \in \mathbb{Z}$. Two loxodromic elements $g, h \in G$ are called *independent* if $\Lambda(g) \cap \Lambda(h) = \varnothing$.

We recall the standard classification of groups acting on hyperbolic spaces, which goes back to Gromov [ibid., Section 8.2].

Theorem 2.5 (Gromov)**.** *For every group G acting on a hyperbolic space S, exactly one of the following conditions holds.*

1) $|\Lambda(G)| = 0$*. Equivalently, G has bounded orbits. In this case the action of G is called* elliptic.

2) $|\Lambda(G)| = 1$*. Equivalently, G has unbounded orbits and contains no loxodromic elements. In this case the action of G is called* parabolic.

3) $|\Lambda(G)| = 2$*. Equivalently, G contains loxodromic elements and any two loxodromic elements have the same limit points. In this case the action of G is called* lineal.

4) $|\Lambda(G)| = \infty$*. Then G always contains loxodromic elements. In turn, this case breaks into two subcases.*

a) *G fixes a point* $\xi \in \partial S$*. In this case* ξ *is the common limit point of all loxodromic elements of* G*. Such an action is called* quasi-parabolic.

b) *G has no fixed points on* ∂S*. Equivalently,* G *contains independent loxodromic elements. In this case the action is said to be of* general type.

Definition 2.6. The action of G is called *elementary* in cases 1)–3) and *non-elementary* in case 4).

An action of a group G on a metric space S is called (metrically) *proper* if the set $\{g \in G \mid \mathrm{d}_S(s, gs) \le r\}$ is finite for all $s \in S$ and $r \in \mathbb{R}_+$. Further, the action of G is *cobounded* if there exists a bounded subset $B \subseteq S$ such that $S = \bigcup_{g \in G} gB$. Finally, the action is *geometric* if it is proper and cobounded. (We work in the category of metric spaces here, so compactness gets replaced by boundedness.)

For geometric actions, we have the following, see Gromov [1987].

Theorem 2.7 (Gromov)**.** *Let* G *be a group acting geometrically on a hyperbolic space. Then exactly one of the following three conditions hold.*

(a) G *acts elliptically. In this case* G *is finite.*

(b) G *acts lineally. In this case* G *is virtually cyclic.*

(c) The action of G *is of general type.*

To every group G generated by a set X one can associate a natural metric space, namely the Cayley graph $\Gamma(G, X)$, on which G acts geometrically. The vertex set of $\Gamma(G, X)$ is G itself and two elements g, h are connected by an edge if $g = hx$ for some $x \in X^{\pm 1}$. This graph is endowed with the *combinatorial metric* induced by identification of edges with $[0, 1]$.

Definition 2.8. A group G is *hyperbolic* if it admits a geometric action on a hyperbolic space.

Equivalently, a group G generated by a finite set X is hyperbolic if the Cayley graph $\Gamma(G, X)$ is a hyperbolic metric space. The equivalence of these two definitions follows from the well-known Svarc-Milnor Lemma and quasi-isometry invariance of hyperbolicity of geodesic spaces, see Bridson and Haefliger [1999] and Gromov [1987] for details.

2.2. Equivalent definitions of acylindrical hyperbolicity. Recall that the action of a group G on a metric space S is *acylindrical* if for every $\varepsilon > 0$ there exist $R, N > 0$ such that for every two points x, y with $\mathrm{d}(x, y) \ge R$, there are at most N elements $g \in G$ satisfying

$$\mathrm{d}(x, gx) \le \varepsilon \quad \text{and} \quad \mathrm{d}(y, gy) \le \varepsilon.$$

The notion of acylindricity goes back to paper Sela [1997], where it was considered for groups acting on trees. In the context of general metric spaces, the above definition is due to Bowditch [2008]. Informally, one can think of this condition as a kind of properness of the action on $S \times S$ minus a "thick diagonal".

Example 2.9. (a) If S is a bounded space, then every action $G \curvearrowright S$ is acylindrical. Indeed it suffices to take $R > \operatorname{diam}(S)$.

(b) It is easy to see that every geometric action is acylindrical. On the other hand, proper actions need not be acylindrical in general.

We begin with a classification of groups acting acylindrically on hyperbolic spaces. The following theorem is proved by Osin [2016] and should be compared to compared to Theorems 2.5 and 2.7.

Theorem 2.10. *Let G be a group acting acylindrically on a hyperbolic space. Then exactly one of the following three conditions holds.*

(a) G acts elliptically, i.e., G has bounded orbits.

(b) G acts lineally. In this case G is virtually cyclic.

(c) The action of G is of general type.

Compared to the general classification of groups acting on hyperbolic spaces, Theorem 2.10 rules out parabolic and quasi-parabolic actions and characterizes groups acting lineally. On the other hand, compared to Theorem 2.7, finiteness of elliptic groups is lacking. This part of Theorem 2.10 cannot be improved, see Example 2.9 (a).

Applying the theorem to cyclic groups, we obtain the following result first proved by Bowditch [2008].

Corollary 2.11. *Every element of a group acting acylindrically on a hyperbolic space is either elliptic or loxodromic.*

Definition 2.12. We call a group G *acylindrically hyperbolic* if it admits a non-elementary acylindrical action on a hyperbolic space. By Theorem 2.10, this is equivalent to the requirement that G is not virtually cyclic and admits an acylindrical action on a hyperbolic space with unbounded orbits.

Unfortunately, Definition 2.12 is hard to verify in practice. Instead, one often first proves that the group satisfies a seemingly weaker condition, which turns out to be equivalent to acylindrical hyperbolicity. To formulate this condition we need a notion introduced by Bestvina and Fujiwara [2002].

Definition 2.13. Let G be a group acting on a hyperbolic space S, g an element of G. One says that g satisfies the *weak proper discontinuity* condition (or g is a *WPD element*) if for every $\varepsilon > 0$ and every $s \in S$, there exists $M \in \mathbb{N}$ such that

$$(1) \qquad \left|\{a \in G \mid \mathrm{d}_S(s, as) < \varepsilon,\ \mathrm{d}(g^M s, ag^M s) < \varepsilon\}\right| < \infty.$$

Obviously this condition holds for any $g \in G$ if the action of G is proper and for every loxodromic $g \in G$ if G acts on S acylindrically.

Theorem 2.14 (Osin [2016, Theorem 1.2]). *For any group G, the following conditions are equivalent.*

(a) G is acylindrically hyperbolic.

(b) G is not virtually cyclic and admits an action on a hyperbolic space such that at least one element of G is loxodromic and satisfies the WPD condition.

(c) There exists a generating set X of G such that the corresponding Cayley graph $\Gamma(G, X)$ is hyperbolic, $|\partial\Gamma(G, X)| > 2$, and the natural action of G on $\Gamma(G, X)$ is acylindrical.

Part (c) of this theorem is especially useful for studying properties of acylindrically hyperbolic groups since it allows to pass from a (possibly non-cobounded) action of G on a general hyperbolic space to the more familiar action on the Cayley graph. It was recently proved by Balasubramanya [2017] that one can ensure even a stronger condition, namely that $\Gamma(G, X)$ is quasi-isometric to a tree.

2.3. Examples. Obviously every geometric action is acylindrical. In particular, this applies to the action of any finitely generated group on its Cayley graph with respect to a finite generating set. Thus every hyperbolic group is virtually cyclic or acylindrically hyperbolic. More generally, non-virtually-cyclic relatively hyperbolic groups with proper peripheral subgroups are acylindrically hyperbolic. In the latter case the action on the relative Cayley graph is non-elementary and acylindrical, see Osin [2016]. Below we discuss some less obvious examples.

(a) *Mapping class groups.* The mapping class group $MCG(\Sigma_{g,p})$ of a closed surface of genus g with p punctures is acylindrically hyperbolic unless $g = 0$ and $p \le 3$ (in these exceptional cases, $MCG(\Sigma_{g,p})$ is finite). For $(g, p) \in \{(0, 4), (1, 0), (1, 1)\}$ this follows from the fact that $MCG(\Sigma_{g,p})$ is non-elementary hyperbolic. For all other values of (g, p) this follows from hyperbolicity of the curve complex $\mathcal{C}(\Sigma_{g,p})$ of $\Sigma_{g,p}$ first proved by Masur and Minsky [1999] and acylindricity of the action of $MCG(\Sigma_{g,p})$ on $\mathcal{C}(\Sigma_{g,p})$, which is due to Bowditch [2008].

(b) $Out(F_n)$. Let $n \geq 2$ and let F_n be the free group of rank n. Bestvina and Feighn [2010] proved that for every fully irreducible automorphism $f \in Out(F_n)$ there exists a hyperbolic graph such that $Out(F_n)$ acts on it and the action of f satisfies the weak proper discontinuity condition. Thus $Out(F_n)$ is acylindrically hyperbolic by Theorem 2.14.

(c) *Groups acting on $CAT(0)$ spaces.* Sisto [2011] showed that if a group G acts properly on a proper $CAT(0)$ space and contains a rank one element, then G is either virtually cyclic or acylindrically hyperbolic. Together with the work of Caprace and Sageev [2011], this implies the following alternative for right angled Artin groups: every right angled Artin group is either cyclic, decomposes as a direct product of two non-trivial groups, or acylindrically hyperbolic. An alternative proof of the later result was found by Kim and Koberda [2014]. A similar theorem holds for graph products of groups and, even more generally, subgroups of graph products, see Minasyan and Osin [2015]. For a survey of examples of acylindrically hyperbolic groups arising from actions on $CAT(0)$ cubical complexes we refer to Genevois [2017].

(d) *Fundamental groups of graphs of groups.* The following theorem was proved by Minasyan and Osin [2015].

Theorem 2.15. *Let G be a group acting minimally on a simplicial tree T. Suppose that G does not fix any point of ∂T and there exist vertices u, v of T such that the pointwise stabilizer of $\{u, v\}$ is finite. Then G is either virtually cyclic or acylindrically hyperbolic.*

If G is the fundamental group of a graph of groups $\mathcal{G}$, then one can apply Theorem 2.15 to the action of G on the associated Bass-Serre tree. In this case the minimality of the action and the absence of fixed points on ∂T can be recognized from the local structure of $\mathcal{G}$. We mention here two particular cases. We say that a subgroup C of a group G is *weakly malnormal* if there exists $g \in G$ such that $|C^g \cap C| < \infty$.

Corollary 2.16. *Let G split as a free product of groups A and B with an amalgamated subgroup C. Suppose $A \neq C \neq B$ and C is weakly malnormal in G. Then G is either virtually cyclic or acylindrically hyperbolic.*

Note that the virtually cyclic case cannot be excluded from this corollary. Indeed it realizes if C is finite and has index 2 in both factors.

Corollary 2.17. *Let G be an HNN-extension of a group A with associated subgroups C and D. Suppose that $C \neq A \neq D$ and C is weakly malnormal in G. Then G is acylindrically hyperbolic.*

These results were used by Minasyan and Osin [ibid.] to prove acylindrical hyperbolicity of a number of groups. E.g., it implies that for every field k, the automorphism group $Aut\, k[x, y]$ of the polynomial algebra $k[x, y]$ is acylindrically hyperbolic. Some other applications are discussed below.

(e) 3*-manifold groups.* Minasyan and Osin [2015] (see also Minasyan and Osin [2017]) proved that for every compact orientable irreducible 3-manifold M, the fundamental group $\pi_1(M)$ is either virtually polycyclic, or acylindrically hyperbolic, or M is Seifert fibered. In the latter case, $\pi_1(M)$ contains a normal subgroup $N \cong \mathbb{Z}$ such that $\pi_1(M)/N$ is acylindrically hyperbolic.

(f) *Groups of deficiency at least* 2*.* In Osin [2015], the author proved that every group which admits a finite presentation with at least 2 more generators than relations is acylindrically hyperbolic. (The original proof contained a gap which is fixed in Minasyan and Osin [2017].) Interestingly, the proof essentially uses results about ℓ^2-Betti numbers of groups.

(g) *Miscellaneous examples.* Other examples include central quotients of Artin-Tits groups of spherical type (see Calvez and Wiest [2017]) and of FC type with underlying Coxeter graph of diameter at least 3 (see Chatterji and Martin [2016]), small cancellation groups, including infinitely presented ones, (see Gruber and Sisto [2014]), orthogonal forms of Kac–Moody groups over arbitrary fields (see Caprace and Hume [2015]), the Cremona group (see Dahmani, Guirardel, and Osin [2017] and references therein; the main contribution towards this result is due to Cantat and Lamy [2013]), and non-elementary convergence groups (see Sun [2017]).

2.4. Some algebraic and analytic properties. Our next goal is to survey some algebraic and analytic properties of acylindrically hyperbolic groups.

(a) *Finite radical.* Every acylindrically hyperbolic group G contains a unique maximal finite normal subgroup denoted $K(G)$ and called the *finite radical* of G, see Dahmani, Guirardel, and Osin [2017]. It also coincides with the amenable radical of G. In particular, G has no infinite amenable normal subgroups.

(b) *SQ-univerality.* Recall that a group G is *SQ-universal* if every countable group can be embedded into a quotient of G. Informally, this property can be considered as an indication of algebraic "largeness" of G. Dahmani, Guirardel, and Osin [ibid.] proved the following result by using group theoretic Dehn filling.

Theorem 2.18. *Every acylindrically hyperbolic group is SQ-universal.*

One consequence of this, also obtained by Dahmani, Guirardel, and Osin [ibid.], is that every subgroup of the mapping class group $MCG(\Sigma)$ of a punctured closed surface Σ is either virtually abelian or SQ-universal. It is easy to show using cardinality arguments that every finitely generated SQ-universal group has uncountably many non-isomorphic quotients. This observation allows one to reprove various (well-known) non-embedding theorems for higher rank lattices in mapping class groups since these lattices have countably many normal subgroups by the Margulis normal subgroup theorem. For instance, we immediately obtain that every homomorphism from an irreducible lattice in a connected semisimple Lie group of $\mathbb{R}$-rank at least 2 with finite center to $MCG(\Sigma)$ has finite image (compare to the main result of Farb and Masur [1998]).

(c) *Mixed identities.* A group G satisfies a *mixed identity* $w = 1$ for some $w \in G * F_n$, where F_n denotes the free group of rank n, if every homomorphism $G * F_n \to G$ that is identical on G sends w to 1. A mixed identity $w = 1$ is non-trivial if $w \neq 1$ as an element of $G * F_n$. We say that G is *mixed identity free* (or *MIF* for brevity) if it does not satisfy any non-trivial mixed identity.

The property of being MIF is much stronger than being identity free and imposes strong restrictions on the algebraic structure of G. For example, if G has a non-trivial center, then it satisfies the non-trivial mixed identity $[a, x] = 1$, where $a \in Z(G) \setminus \{1\}$. Similarly, it is easy to show (see Hull and Osin [2016b]) that a MIF group has no finite normal subgroups, is directly indecomposable, has infinite girth, etc. By constructing highly transitive permutation representations of acylindrically hyperbolic groups, Hull and the author proved that every acylindrically hyperbolic group with trivial finite radical is MIF.

(d) *Quasi-cocycles and bounded cohomology.* The following theorem was proved in several papers under various assumptions (see Bestvina, Bromberg, and Fujiwara [2016], Bestvina and Fujiwara [2002], Hamenstädt [2008], and Hull and Osin [2013] and references therein), which later turned out to be equivalent to acylindrical hyperbolicity.

Theorem 2.19. *Suppose that a group G is acylindrically hyperbolic. Let $V = \mathbb{R}$ or $V = \ell^p(G)$ for some $p \in [1, +\infty)$. Then the kernel of the natural map $H^2_b(G, V) \to H^2(G, V)$ is infinite dimensional. In particular,* $\dim H^2_b(G, V) = \infty$.

This result opens the door for Monod and Shalom [2006] rigidity theory for group actions on spaces with measure. It also implies that acylindrically hyperbolic groups are not boundedly generated, i.e., are not products of finitely many cyclic subgroups.

(e) *Stability properties.* It is not difficult to show that the class of acylindrically hyperbolic groups is stable under taking extensions with finite kernel and quotients modulo finite normal subgroups. It is also stable under taking finite index subgroups and, more generally, s-normal subgroups Osin [2016]. Recall that a subgroup N of a group G is s-normal if $g^{-1}Ng \cap N$ is infinite for all $g \in G$.

On the other hand, it is not known if acylindrical hyperbolicity is stable under finite extensions (see Minasyan and Osin [2017]). More generally, we propose the following.

Question 2.20. *(a) Is acylindrical hyperbolicity of finitely generated groups a quasi-isometry invariant?*

(b) Is acylindrical hyperbolicity a measure equivalence invariant?

The last question is partially motivated by the fact that the property $H^2_b(G, \ell^2(G)) \neq 0$ enjoyed by all acylyndrically hyperbolic groups by Theorem 2.19 is a measure equivalence invariant. For details we refer to Monod and Shalom [2006].

3 Hyperbolically embedded subgroups

3.1. Definition and basic examples. Hyperbolically embedded collections of subgroups were introduced by Dahmani, Guirardel, and Osin [2017] as generalizations of peripheral subgroups of relatively hyperbolic groups. To simplify our exposition we restrict here to the case of a single subgroup; the general case only differs by notation.

Let G be a group, H a subgroup of G. Suppose that X is a relative generating set of G with respect to H, i.e., $G = \langle X \cup H \rangle$. We denote by $\Gamma(G, X \sqcup H)$ the Cayley graph of G whose edges are labeled by letters from the alphabet $X \sqcup H$. That is, two vertices $f, g \in G$ are connected by an edge going from f to g and labeled by $a \in X \sqcup H$ iff $fa = g$ in G. Disjointness of the union in this definition means that if a letter $h \in H$ and a letter $x \in X$ represent the same element $a \in G$, then for every $g \in G$, the Cayley graph $\Gamma(G, X \sqcup H)$ will have two edges connecting g and ga: one labelled by h and the other labelled by x.

We naturally think of the Cayley graph $\Gamma_H = \Gamma(H, H)$ of H with respect to the generating set H as a (complete) subgraph of $\Gamma(G, X \sqcup H)$.

Definition 3.1. Let G be a group, $H \leq G$, and X a (possibly infinite) subset of G. We say that H is *hyperbolically embedded in G with respect to X* (we write $H \hookrightarrow_h (G, X)$) if $G = \langle X \cup H \rangle$ and the following conditions hold.

(a) The Cayley graph $\Gamma(G, X \sqcup H)$ is hyperbolic.

(b) For every $n \in \mathbb{N}$, there are only finitely many elements $h \in H$ such that the vertices h and 1 can be connected in $\Gamma(G, X \sqcup H)$ by a path of length at most n that avoids edges of Γ_H.

Further we say that H is hyperbolically embedded in G and write $H \hookrightarrow_h G$ if $H \hookrightarrow_h (G, X)$ for some $X \subseteq G$.

Note that for any group G we have $G \hookrightarrow_h G$. Indeed we can take $X = \varnothing$ in this case. Further, if H is a finite subgroup of a group G, then $H \hookrightarrow_h G$. Indeed $H \hookrightarrow_h (G, X)$ for $X = G$. These cases are referred to as *degenerate*. We consider two additional examples borrowed from Dahmani, Guirardel, and Osin [ibid.].

Example 3.2. (a) Let $G = H \times \mathbb{Z}$, $X = \{x\}$, where x is a generator of $\mathbb{Z}$. Then $\Gamma(G, X \sqcup H)$ is quasi-isometric to a line and hence it is hyperbolic. However, every two elements $h_1, h_2 \in H$ can be connected by a path of length at most 3 in $\Gamma(G, X \sqcup H)$ that avoids edges of Γ_H (see Figure 1). Thus $H \not\hookrightarrow_h (G, X)$ whenever H is infinite.

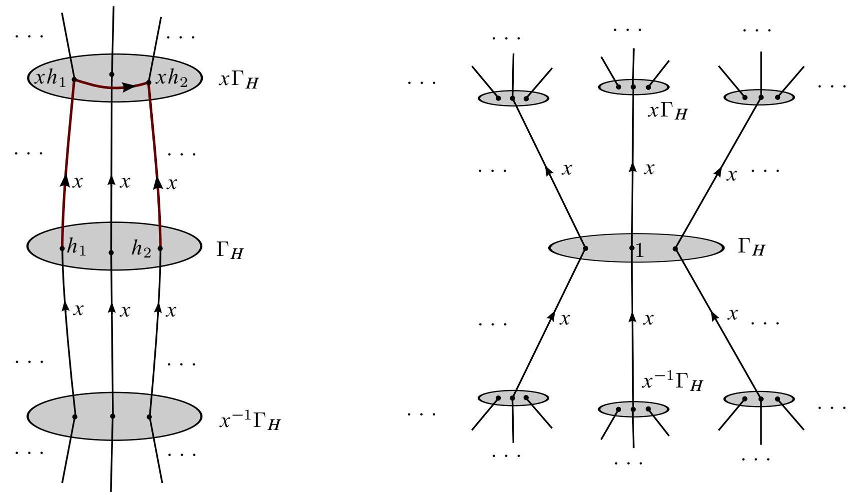


Figure 1: Cayley graphs $\Gamma(G, X \sqcup H)$ for $G = H \times \mathbb{Z}$ and $G = H * \mathbb{Z}$.

(b) Let $G = H * \mathbb{Z}$, $X = \{x\}$, where x is a generator of $\mathbb{Z}$. In this case $\Gamma(G, X \sqcup H)$ is quasi-isometric to a tree and no path connecting $h_1, h_2 \in H$ and avoiding edges of Γ_H exists unless $h_1 = h_2$. Thus $H \hookrightarrow_h (G, X)$.

It is worth noting that a version of the argument from Example 3.2 (a) can be used to show that every hyperbolically embedded subgroup $H \hookrightarrow_h G$ is *almost malnormal*, i.e., satisfies $|g^{-1}Hg \cap H| < \infty$ for all $g \in G$.

The following result is proved by Dahmani, Guirardel, and Osin [ibid.] and can be regarded as a definition of relatively hyperbolic groups.

Theorem 3.3. *Let G be a group, H a subgroup of G. Then G is hyperbolic relative to H if and only if $H \hookrightarrow_h (G, X)$ for some finite subset $X \subseteq G$.*

3.2. Hyperbolically embedded subgroups in acylindrically hyperbolic groups. It turns out that acylindrical hyperbolicity of a group can be characterized by the existence of hyperbolically embedded subgroups. More precisely, we have the following.

Theorem 3.4 (Osin [2016]). *A group G is acylindrically hyperbolic if and only if it contains non-degenerate hyperbolically embedded subgroups.*

Moreover, in every acylindrically hyperbolic group one can find hyperbolically embedded subgroups of certain special types. We mention two results of this sort proved by Dahmani, Guirardel, and Osin [2017]. The first one plays an important role in applications of group theoretic Dehn filling and small cancellation theory discussed below.

Theorem 3.5. *Let G be a group acting on a hyperbolic space and let $g \in G$ be a loxodromic WPD element. Then g is contained in a unique maximal virtually cyclic subgroup $E(g)$ of G and $E(g) \hookrightarrow_h G$.*

Recall that $K(G)$ denotes the final radical of an acylindrically hyperbolic group G and F_n denotes the free group of rank n.

Theorem 3.6. *Let G be an acylindrically hyperbolic group. Then for every $n \in \mathbb{N}$, there exists a subgroup $H \hookrightarrow_h G$ isomorphic to $F_n \times K(G)$.*

The latter theorem is especially useful in conjunction with various "extension" results proved by Abbott, Hume, and Osin [2017], Frigerio, Pozzetti, and Sisto [2015], and Hull and Osin [2013]. Roughly speaking, these results claim that various things (e.g., group actions on metric spaces or quasi-cocycles) can be "extended" from a hyperbolically embedded subgroup to the whole group.

4 Group theoretic Dehn filling

4.1. Dehn surgery in 3-manifolds. Dehn surgery on a 3-dimensional manifold consists of cutting of a solid torus from the manifold, which may be thought of as "drilling" along an embedded knot, and then gluing it back in a different way. The study of these "elementary transformations" of 3-manifolds is partially motivated by the Lickorish-Wallace theorem, which states that every closed orientable connected 3-manifold can be obtained by performing finitely many surgeries on the 3-dimensional sphere.

The second part of the surgery, called *Dehn filling*, can be formalized as follows. Let M be a compact orientable 3-manifold with toric boundary. Topologically distinct ways to attach a solid torus to ∂M are parameterized by free homotopy classes of unoriented essential simple closed curves in ∂M, called *slopes*. For a slope s, the corresponding Dehn filling $M(s)$ of M is the manifold obtained from M by attaching a solid torus $\mathbb{D}^2 \times \mathbb{S}^1$ to ∂M so that the meridian $\partial \mathbb{D}^2$ goes to a simple closed curve of the slope s.

The following fundamental theorem is due to Thurston [1982, Theorem 1.6].

Theorem 4.1 (Thurston's hyperbolic Dehn surgery theorem)**.** *Let M be a compact orientable 3-manifold with toric boundary. Suppose that $M \setminus \partial M$ admits a complete finite volume hyperbolic structure. Then $M(s)$ is hyperbolic for all but finitely many slopes s.*

4.2. Filling in hyperbolically embedded subgroups. Dehn filling can be generalized in the context of abstract group theory as follows. Let G be a group and let H be a subgroup of G. One can think of G and H as the analogues of $\pi_1(M)$ and $\pi_1(\partial M)$, respectively. Associated to any $\sigma \in H$, is the quotient group $G/\langle\langle s \rangle\rangle$, where $\langle\langle s \rangle\rangle$ denotes the normal closure of s in G.

If $G = \pi_1(M)$ and $H = \pi_1(\partial M) \cong \mathbb{Z} \oplus \mathbb{Z}$, where M is as in Thurston's theorem, then H is indeed a subgroup of G and for every slope s, which we think of as an element of H, we have

$$\pi_1(M(s)) = \pi_1(M)/\langle\langle s\rangle\rangle \tag{2}$$

by the Seifert-van Kampen theorem. Thus $G/\langle\langle s\rangle\rangle$ is the algebraic counterpart of the filling $M(s)$.

It turns out that the analogue of Thurston's theorem holds if we start with a pair $H \leq G$ such that H is hyperbolically embedded in G. The vocabulary translating geometric terms to algebraic ones can be summarized as follows (we abbreviate "complete finite volume" as CFV).

3-MANIFOLDS	**GROUPS**
a compact orientable 3-manifold M	a group G
∂M	$H \leq G$
$M \setminus \partial M$ admits a finite volume hyperbolic structure	H is hyperbolically embedded in G
a slope s	an element $h \in H$
$M(s)$	$G/\langle\langle h\rangle\rangle$

In these settings, the analogue of Thurston's theorem was proved by Dahmani, Guirardel, and Osin [2017]. Note that instead of considering single elements of H, we allow normal subgroups generated by arbitrary sets of elements. A number of additional properties can be added to the main statements (a)–(c); we mention just one of them, which is necessary for the applications considered in the next section.

Theorem 4.2. *Let G be a group, H a subgroup of G. Suppose that $H \hookrightarrow_h (G, X)$ for some $X \subseteq G$. Then there exists a finite subset $\mathfrak{F}$ of nontrivial elements of H such that for every subgroup $N \lhd H$ that does not contain elements of $\mathfrak{F}$, the following hold.*

(a) If G is acylindrically hyperbolic, then so is $G/\langle\langle N\rangle\rangle$, where $\langle\langle N\rangle\rangle$ denotes the normal closure of N in G.

(b) The natural map from H/N to $G/\langle\langle N\rangle\rangle$ is injective (equivalently, $H \cap \langle\langle N\rangle\rangle = N$).

(c) $H/N \hookrightarrow_h (G/\langle\langle N\rangle\rangle, \overline{X})$*, where* $\overline{X}$ *is the natural image of* X *in* $G/\langle\langle N\rangle\rangle$*.*

(d) $\langle\langle N\rangle\rangle$ *is the free product of conjugates of* N *in* G *and every element of* $\langle\langle N\rangle\rangle$ *is either conjugate to an element of* N *or acts loxodromically on* $\Gamma(G, X \sqcup H)$*.*

Note that if $H \hookrightarrow_h G$ is non-degenerate, then G is always acylindrically hyperbolic. However the theorem holds (trivially) for degenerate hyperbolically embedded subgroups as well.

Combining this theorem with Theorem 3.3 and some basic properties of relatively hyperbolic groups, we obtain the following result, which was first proved by Osin [2007]. It was also independently proved by Groves and Manning [2008] under the additional assumptions that the group G is torsion free and finitely generated.

Corollary 4.3. *Suppose that a group* G *is hyperbolic relative to a subgroup* $H \neq G$*. Then for any subgroup* $N \lhd H$ *avoiding a fixed finite set of nontrivial elements, the natural map from* H/N *to* $G/\langle\langle N\rangle\rangle$ *is injective and* $G/\langle\langle N\rangle\rangle$ *is hyperbolic relative to* H/N*. In particular, if* H/N *is hyperbolic, then so is* $G/\langle\langle N\rangle\rangle$*; if, in addition,* G *is non-virtually-cyclic, then so is* $G/\langle\langle N\rangle\rangle$*.*

Under the assumptions of Thurston's theorem, we have $H = \pi_1(\partial M) = \mathbb{Z} \oplus \mathbb{Z}$. Slopes in ∂M correspond to non-trivial primitive elements $s \in H$; for every such s, we have $H/\langle s\rangle \cong \mathbb{Z}$. Applying Corollary 4.3 to $N = \langle s\rangle \lhd H$, we obtain that $G/\langle\langle N\rangle\rangle$ is not virtually cyclic and hyperbolic. Modulo the geometrization conjecture this algebraic statement is equivalent to hyperbolicity of $M(s)$. Thus parts (a)–(c) of Theorem 4.2 indeed provide a group theoretic generalization of Thurston's theorem.

4.3. Applications. It is not feasible to discuss all applications of group theoretic Dehn surgery in a short survey. Here we list some of the results which make use of Theorem 4.2 or its relatively hyperbolic analogue, Corollary 4.3, and provide references for further reading. Then we pick one application and discuss it in more detail.

(a) *The virtual Haken conjecture.* Group theoretic Dehn filling in relatively hyperbolic groups, along with Wise's machinery of virtually special groups, was used in Agol's proof of the virtual Haken conjecture Agol [2013]. Additional results on Dehn filling necessary for the proof were obtained by Agol, Groves, and Manning in the appendix to Agol [ibid.]. One piece of Wise's work used by Agol [ibid.] is the malnormal special quotient theorem; Agol, Groves, and Manning [2016] also found an alternative proof of this result based on Dehn filling technique.

(b)*The isomorphism problem for relatively hyperbolic groups.* Dahmani and Guirardel [2015] and Dahmani and Touikan [2013] used Dehn filling to solve the isomorphism problem for relatively hyperbolic groups with residually finite parabolic subgroups under certain additional assumptions. The main idea is to apply (an elaborated version of) Corollary 4.3 to finite index normal subgroups in parabolic groups. This yields an approximation of relatively hyperbolic groups by hyperbolic ones, which in turn allows the authors make use of the solution of the isomorphism problem for hyperbolic groups obtained by Dahmani and Guirardel [2011].

(c) *Residual finiteness of outer automorphism groups.* Minasyan and Osin [2010] used Dehn filling in relatively hyperbolic groups to prove that $Out(G)$ is residually finite for every residually finite group G with infinitely many ends; in general, this result fails for one ended groups. This result was recently generalized to conjugacy separable acylindrically hyperbolic groups by Antolin, Minasyan, and Sisto. In particular, they proved residual finiteness of mapping class groups of certain Haken 3-manifolds. Acylindrical hyperbolicity of 3-manifold groups plays a crucial role in the proof.

(d) *Primeness of von Neumann algebras.* Chifan, Kida, and Pant [2016] used Dehn filling to prove primeness of von Neumann algebras of certain relatively hyperbolic groups. This means that these von Neumann algebras cannot be decomposed as a tensor product of diffuse von Neumann algebras.

(e) *Farell-Jones conjecture for relatively hyperbolic groups.* Bartels [2017] proved that the class of groups satisfying the Farell-Jones conjecture is stable under relative hyperbolicity. In the particular case when peripheral subgroups are residually finite, an alternative proof based on Dehn filling was found by Antolín, Coulon, and Gandini [2015].

(f) *SQ-universality of acylindrically hyperbolic groups.* One simple application of Theorem 4.2 is the proof of Theorem 2.18. It follows easily from SQ-universality of free groups of rank 2, Theorem 3.6, and part (b) of Theorem 4.2. For details, see Dahmani, Guirardel, and Osin [2017].

4.4. Purely pseudo-Anosov subgroups of mapping class groups. We illustrate Theorem 4.2 by considering an application to mapping class groups. Recall that a subgroup of a mapping class group is called *purely pseudo-Anosov*, if all its non-trivial elements are pseudo-Anosov. The following question is Problem 2.12(A) in Kirby's list: *Does the mapping class group of any closed orientable surface of genus* $g \geq 1$ *contain a non-trivial purely pseudo-Anosov normal subgroup?* It was asked in the early 1980s and is often attributed to Penner, Long, and McCarthy. It is also recorded by Ivanov [2006, Problems 3], and Farb [2006] refers to it as a "well known open question".

The abundance of finitely generated non-normal purely pseudo-Anosov free subgroups of mapping class groups is well known, and follows from an easy ping-pong argument.

However, this method does not allow one to construct normal subgroups, which are usually infinitely generated. For a surface of genus 2 the question was answered by Whittlesey [2000] who proposed an example based on Brunnian braids. Unfortunately the methods of Whittlesey [ibid.] do not generalize even to closed surfaces of higher genus.

Another question was probably first asked by Ivanov (see Ivanov [2006, Problem 11]): *Is the normal closure of a certain nontrivial power of a pseudo-Anosov element of $MCG(S_g)$ free?* Farb [2006, Problem 2.9] also described this problem as a "basic test question" for understanding normal subgroups of mapping class groups.

The next theorem answers both questions positively in more general settings.

Theorem 4.4 (Theorem 2.30, Dahmani, Guirardel, and Osin [2017]). *Let G be a group acting on a hyperbolic space S, $g \in G$ a WPD loxodromic element. Then there exists $n \in \mathbb{N}$ such that the normal closure $\langle\langle g^n \rangle\rangle$ in G is free and purely loxodromic, i.e., every nontrivial element of $\langle\langle g^n \rangle\rangle$ acts loxodromically on S.*

This result can be viewed as a generalization of a theorem by Delzant [1996] stating that for a hyperbolic group G and every element of infinite order $g \in G$, there exists $n \in \mathbb{N}$ such that $\langle\langle g^n \rangle\rangle$ is free (see also Chaynikov [2011] for a clarification of certain aspects of Delzant's proof).

The idea of the proof is the following. By Theorem 3.5, g is contained in the maximal virtually cyclic subgroup $E(g)$ which is hyperbolically embedded in G. Since $\langle g \rangle$ has finite index in $E(g)$, we have $\langle g^n \rangle \lhd E(g)$. Passing to a multiple of n if necessary, we can ensure that $\langle g^n \rangle$ avoids any finite collection of non-trivial elements. Thus we can apply Theorem 4.2 to $H = E(g)$ and $N = \langle g^n \rangle$. Since $\langle g^n \rangle \cong \mathbb{Z}$, part (d) of the theorem implies that $\langle\langle g^n \rangle\rangle$ is free. That $\langle\langle g^n \rangle\rangle$ is purely loxodromic also follows from part (d) and some additional arguments relating $\Gamma(G, X \sqcup H)$ to S.

Applying Theorem 4.4 to mapping class groups acting on the curve complexes, we obtain the following.

Corollary 4.5. *Let Σ be a possibly punctured closed orientable surface. Then for any pseudo-Anosov element $a \in MCG(\Sigma)$, there exists $n \in \mathbb{N}$ such that the normal closure of a^n is free and purely pseudo-Anosov.*

5 Small cancellation theory and its applications

5.1. Generalizing classical small cancellation. The classical small cancellation theory deals with presentations

$$F(X)/\langle\langle \mathcal{R} \rangle\rangle = \langle X \mid \mathcal{R} \rangle,$$

where $F(X)$ is the free group with basis X, and common subwords of distinct relators are "small" in a certain precise sense. This property allows one to control cancellation in products of conjugates of relators in $\mathcal{R}$ (and their inverses); in turn, this leads to nice structural results for the normal closure $\langle\langle \mathcal{R} \rangle\rangle$ and the group $F(X)/\langle\langle \mathcal{R} \rangle\rangle$.

More generally, one can replace the free group $F(X)$ with a group G_0 enjoying some hyperbolic properties and add new relations to a presentation of G_0. If these new relations satisfy a suitable version of small cancellation, many results of the classical small cancellation theory can be reproved in these settings. On the other hand, the small cancellation assumptions are usually general enough to allow one to create interesting relations between elements.

The idea of generalizing classical small cancellation to groups acting on hyperbolic spaces is due to Gromov [1987], although some underlying ideas go back to the work of Olshanskii [1982, 1980]. In the case of hyperbolic groups, it was formalized by Delzant [1996], Olshanskii [1993], and others. Olshanskii's approach was generalized to relatively hyperbolic groups by Osin [2010] and further generalized to acylindrically hyperbolic groups by Hull [2016]. These generalizations employ isoperimetric characterizations of relatively hyperbolic groups and hyperbolically embedded subgroups obtained in Dahmani, Guirardel, and Osin [2017] and Osin [2006] and follow closely the classical theory. Yet another approach is based on Gromov's *rotating families* (see Coulon [2016] and references therein.)

Unfortunately, the ideas involved in this work are too technical for a short survey paper and we do not discuss them here. Instead we discuss few (indeed very few) applications of small cancellation theory in relatively hyperbolic groups to proving embedding theorems and studying conjugacy growth of groups.

5.2. Embedding theorems and conjugacy growth of groups. In 1949, Higman, B. H. Neumann, and H. Neumann [1949] proved that any countable group G can be embedded into a countable group B such that every two elements of the same order are conjugate in B. The group B constructed by Higman, B. H. Neumann, and H. Neumann [ibid.] is a union of infinite number of subsequent HNN–extensions and thus it is never finitely generated. Osin [2010] used small cancellation theory in relatively hyperbolic groups to prove the following stronger result. For a group G, let $\pi(G)$ denote the set of finite orders of elements of G.

Theorem 5.1. *Any countable group G can be embedded into a finitely generated group C such that any two elements of the same order are conjugate in C and $\pi(G) = \pi(C)$.*

We explain the idea of the proof in the particular case when C is torsion free. Let $G_0 = C * F(x, y)$, where $F(x, y)$ is the free group with basis $\{x, y\}$. Given any nontrivial lement $g \in G_0$, one first considers the HNN-extension

$$H = \langle G_0, t \mid t^{-1}gt = x\rangle.$$

Obviously x and g are conjugate in H. Then imposing an additional relation $t = w(x, y)$, where $w(x, y)$ is a suitable small cancellation word in the alphabet $\{x, y\}$, one ensures that this conjugation happens in a certain quotient group G_1 of G_0. Small cancellation theory is then used to show that the restriction of the natural homomorphism $G_0 \to G_1$ to C is injective and the image of $F(x, y)$ in G_1 is still "large enough". Here "large enough" means that the image of $F(x, y)$ in G_1 is non-elementary with respect to some acylindrical action of G_1 on a hyperbolic space. This allows us to iterate the process. Repeating it for all non-trivial elements we obtain a group with 2 conjugacy classes which is generated by 2 elements (the images of x and y) and contains C.

Applying the theorem to the group $G = \mathbb{Z}$, we obtain the following.

Corollary 5.2. *There exists a torsion free finitely generated group with* 2 *conjugacy classes.*

The existence of a finitely generated group with 2 conjugacy classes other than $\mathbb{Z}/2\mathbb{Z}$ was a long standing open problem, sometimes attributed to Maltsev. It is easy to see that such groups do not exist among finite (and residually finite) groups. It is also observed by Osin [2010] that such a group cannot be constructed as a limit of hyperbolic groups; this justifies the use of small cancellation theory in the more general settings.

Given a group G generated by a finite set X, the associated *conjugacy growth function* of G, denoted by $\xi_{G,X}$, is defined as follows: $\xi_{G,X}(n)$ is the number of conjugacy classes of elements that can be represented by words of length at most n is the alphabet $X \cup X^{-1}$. Given $f, g\colon \mathbb{N} \to \mathbb{N}$, we write $f \sim g$ if there exists $C \in \mathbb{N}$ such that $f(n) \le g(Cn)$ and $g(n) \le f(Cn)$ for all $n \in \mathbb{N}$. Obviously $\sim$ is an equivalence relation and $\xi_{G,X}(n)$ is independent of the choice of X up to this equivalence.

The conjugacy growth function was introduced by Babenko [1988] in order to study geodesic growth of Riemannian manifolds. For more details and a survey of some recent results about conjugacy growth we refer to Hull and Osin [2016a]. Based on ideas from the paper Osin [2010], Hull and the author also obtained a complete description of functions that occur as conjugacy growth functions of finitely generated groups. It is worth noting that such a description for the usual growth function seems to be out of reach at this time.

Theorem 5.3. *Let G be a group generated by a finite set X, and let f denote the conjugacy growth function of G with respect to X. Then the following conditions hold.*

(a) f is non-decreasing.

(b) There exists $a \ge 1$ such that $f(n) \le a^n$ for every $n \in \mathbb{N}$.

Conversely, suppose that a function $f : \mathbb{N} \to \mathbb{N}$ satisfies the above conditions (a) and (b). Then there exists an group G generated by a finite set X such that $\xi_{G,X} \sim f$.

Of course, the non-trivial part of the theorem is the fact that every function satisfying (a) and (b) realizes as the conjugacy growth function.

Yet another result proved by Hull and Osin [2016a] is the following.

Theorem 5.4. *There exists a finitely generated group G and a finite index subgroup $H \leq G$ such that H has 2 conjugacy classes while G is of exponential conjugacy growth.*

In particular, unlike the usual growth function, conjugacy growth of a group is not a quasi-isometry invariant.

Readers interested in other applications of small cancellation technique to groups with hyperbolically embedded subgroups are referred to Hull [2016] and Minasyan and Osin [2018]; for a slightly different approach employing rotating families see Gromov's paper Gromov [2003], Coulon's survey Coulon [2016], and references therein.

References

Carolyn Abbott, David Hume, and Denis Osin (Mar. 2017). "Extending group actions on metric spaces". arXiv: 1703.03010 (cit. on p. 948).

Ian Agol (2013). "The virtual Haken conjecture". *Doc. Math.* 18. With an appendix by Agol, Daniel Groves, and Jason Manning, pp. 1045–1087. MR: 3104553 (cit. on pp. 938, 950).

Ian Agol, Daniel Groves, and Jason Fox Manning (2016). "An alternate proof of Wise's malnormal special quotient theorem". *Forum Math. Pi* 4, e1, 54. MR: 3456181 (cit. on p. 950).

Yago Antolín, Rémi Coulon, and Giovanni Gandini (Oct. 2015). "Farrell-Jones via Dehn fillings". arXiv: 1510.08113 (cit. on p. 951).

Goulnara Arzhantseva, Ashot Minasyan, and Denis Osin (2007). "The SQ-universality and residual properties of relatively hyperbolic groups". *J. Algebra* 315.1, pp. 165–177. MR: 2344339.

I. K. Babenko (1988). "Closed geodesics, asymptotic volume and the characteristics of growth of groups". *Izv. Akad. Nauk SSSR Ser. Mat.* 52.4, pp. 675–711, 895. MR: 966980 (cit. on p. 954).

Sahana H. Balasubramanya (2017). "Acylindrical group actions on quasi-trees". *Algebr. Geom. Topol.* 17.4, pp. 2145–2176. arXiv: 1602.03941. MR: 3685605 (cit. on p. 942).

A. Bartels (2017). "Coarse flow spaces for relatively hyperbolic groups". *Compos. Math.* 153.4, pp. 745–779. MR: 3631229 (cit. on p. 951).

Mladen Bestvina, Ken Bromberg, and Koji Fujiwara (2016). "Bounded cohomology with coefficients in uniformly convex Banach spaces". *Comment. Math. Helv.* 91.2, pp. 203–218. MR: 3493369 (cit. on p. 945).

Mladen Bestvina and Mark Feighn (2010). "A hyperbolic Out(F_n)-complex". *Groups Geom. Dyn.* 4.1, pp. 31–58. MR: 2566300 (cit. on p. 943).
Mladen Bestvina and Koji Fujiwara (2002). "Bounded cohomology of subgroups of mapping class groups". *Geom. Topol.* 6, pp. 69–89. MR: 1914565 (cit. on pp. 941, 945).
Brian H. Bowditch (2008). "Tight geodesics in the curve complex". *Invent. Math.* 171.2, pp. 281–300. MR: 2367021 (cit. on pp. 941, 942).
Martin R. Bridson and André Haefliger (1999). *Metric spaces of non-positive curvature*. Vol. 319. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, pp. xxii+643. MR: 1744486 (cit. on pp. 938, 940).
Matthieu Calvez and Bert Wiest (2017). "Acylindrical hyperbolicity and Artin-Tits groups of spherical type". *Geom. Dedicata* 191, pp. 199–215. arXiv: 1606.07778. MR: 3719080 (cit. on p. 944).
Serge Cantat and Stéphane Lamy (2013). "Normal subgroups in the Cremona group". *Acta Math.* 210.1. With an appendix by Yves de Cornulier, pp. 31–94. MR: 3037611 (cit. on p. 944).
Pierre-Emmanuel Caprace and David Hume (2015). "Orthogonal forms of Kac-Moody groups are acylindrically hyperbolic". *Ann. Inst. Fourier (Grenoble)* 65.6, pp. 2613–2640. arXiv: 1408.6117. MR: 3449592 (cit. on p. 944).
Pierre-Emmanuel Caprace and Michah Sageev (2011). "Rank rigidity for CAT(0) cube complexes". *Geom. Funct. Anal.* 21.4, pp. 851–891. MR: 2827012 (cit. on p. 943).
Indira Chatterji and Alexandre Martin (Oct. 2016). "A note on the acylindrical hyperbolicity of groups acting on CAT(0) cube complexes". arXiv: 1610.06864 (cit. on p. 944).
Vladimir Chaynikov (2011). "On the generators of the kernels of hyperbolic group presentations". *Algebra Discrete Math.* 11.2, pp. 18–50. MR: 2858127 (cit. on p. 952).
Ionut Chifan, Yoshikata Kida, and Sujan Pant (2016). "Primeness results for von Neumann algebras associated with surface braid groups". *Int. Math. Res. Not. IMRN* 16, pp. 4807–4848. MR: 3556425 (cit. on p. 951).
Rémi Coulon (2016). "Théorie de la petite simplification: une approche géométrique [d'après F. Dahmani, V. Guirardel, D. Osin et S. Cantat, S. Lamy]". *Astérisque* 380, Séminaire Bourbaki. Vol. 2014/2015, Exp. No. 1089, 1–33. MR: 3522169 (cit. on pp. 953, 955).
François Dahmani and Vincent Guirardel (2011). "The isomorphism problem for all hyperbolic groups". *Geom. Funct. Anal.* 21.2, pp. 223–300. MR: 2795509 (cit. on p. 951).
– (June 2015). "Recognizing a relatively hyperbolic group by its Dehn fillings". arXiv: 1506.03233 (cit. on p. 951).
Francois Dahmani, Vincent Guirardel, and Denis Osin (2017). "Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces". *Mem. Amer. Math. Soc.* 245.1156, pp. v+152. MR: 3589159 (cit. on pp. 938, 944, 946, 947, 949, 951–953).

François Dahmani and Nicholas Touikan (Nov. 2013). "Deciding Isomorphy using Dehn fillings, the splitting case". arXiv: 1311.3937 (cit. on p. 951).

Thomas Delzant (1996). "Sous-groupes distingués et quotients des groupes hyperboliques". *Duke Math. J.* 83.3, pp. 661–682. MR: 1390660 (cit. on pp. 952, 953).

Benson Farb (2006). "Some problems on mapping class groups and moduli space". In: *Problems on mapping class groups and related topics*. Vol. 74. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 11–55. MR: 2264130 (cit. on pp. 951, 952).

Benson Farb and Howard Masur (1998). "Superrigidity and mapping class groups". *Topology* 37.6, pp. 1169–1176. MR: 1632912 (cit. on p. 945).

R. Frigerio, M. B. Pozzetti, and A. Sisto (2015). "Extending higher-dimensional quasi-cocycles". *J. Topol.* 8.4, pp. 1123–1155. MR: 3431671 (cit. on p. 948).

Anthony Genevois (Sept. 2017). "Hyperbolicities in CAT(0) cube complexes". arXiv: 1709.08843 (cit. on p. 943).

Étienne Ghys and Pierre de la Harpe (1990). "Espaces métriques hyperboliques". In: *Sur les groupes hyperboliques d'après Mikhael Gromov (Bern, 1988)*. Vol. 83. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 27–45. MR: 1086650 (cit. on p. 938).

M. Gromov (1987). "Hyperbolic groups". In: *Essays in group theory*. Vol. 8. Math. Sci. Res. Inst. Publ. Springer, New York, pp. 75–263. MR: 919829 (cit. on pp. 937–940, 953).

– (2003). "Random walk in random groups". *Geom. Funct. Anal.* 13.1, pp. 73–146. MR: 1978492 (cit. on p. 955).

Daniel Groves and Jason Fox Manning (2008). "Dehn filling in relatively hyperbolic groups". *Israel J. Math.* 168, pp. 317–429. MR: 2448064 (cit. on pp. 938, 950).

Dominik Gruber and Alessandro Sisto (Aug. 2014). "Infinitely presented graphical small cancellation groups are acylindrically hyperbolic". arXiv: 1408.4488 (cit. on p. 944).

Ursula Hamenstädt (2008). "Bounded cohomology and isometry groups of hyperbolic spaces". *J. Eur. Math. Soc. (JEMS)* 10.2, pp. 315–349. MR: 2390326 (cit. on p. 945).

Graham Higman, B. H. Neumann, and Hanna Neumann (1949). "Embedding theorems for groups". *J. London Math. Soc.* 24, pp. 247–254. MR: 0032641 (cit. on p. 953).

Michael Hull (2016). "Small cancellation in acylindrically hyperbolic groups". *Groups Geom. Dyn.* 10.4, pp. 1077–1119. MR: 3605028 (cit. on pp. 938, 953, 955).

Michael Hull and Denis Osin (2013). "Induced quasicocycles on groups with hyperbolically embedded subgroups". *Algebr. Geom. Topol.* 13.5, pp. 2635–2665. MR: 3116299 (cit. on pp. 945, 948).

– (2016a). "Corrigendum to "Conjugacy growth of finitely generated groups" [Adv. Math. 235 (2013) 361–389] [3010062]". *Adv. Math.* 294, pp. 857–859. MR: 3479580 (cit. on pp. 954, 955).

Michael Hull and Denis Osin (2016b). "Transitivity degrees of countable groups and acylindrical hyperbolicity". *Israel J. Math.* 216.1, pp. 307–353. MR: 3556970 (cit. on p. 945).

Nikolai V. Ivanov (2006). "Fifteen problems about the mapping class groups". In: *Problems on mapping class groups and related topics*. Vol. 74. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 71–80. MR: 2264532 (cit. on pp. 951, 952).

Sang-Hyun Kim and Thomas Koberda (2014). "The geometry of the curve graph of a right-angled Artin group". *Internat. J. Algebra Comput.* 24.2, pp. 121–169. MR: 3192368 (cit. on p. 943).

Howard Masur and Yair N. Minsky (1999). "Geometry of the complex of curves. I. Hyperbolicity". *Invent. Math.* 138.1, pp. 103–149. MR: 1714338 (cit. on p. 942).

Ashot Minasyan and Denis Osin (2010). "Normal automorphisms of relatively hyperbolic groups". *Trans. Amer. Math. Soc.* 362.11, pp. 6079–6103. MR: 2661509 (cit. on p. 951).

– (2015). "Acylindrical hyperbolicity of groups acting on trees". *Math. Ann.* 362.3-4, pp. 1055–1105. MR: 3368093 (cit. on pp. 943, 944).

– (Nov. 2017). "Erratum to the paper *Acylindrical hyperbolicity of groups acting on trees*". arXiv: 1711.09486 (cit. on pp. 944, 945).

– (Apr. 2018). "Acylindrically hyperbolic groups with exotic properties". arXiv: 1804.08767 (cit. on p. 955).

Nicolas Monod and Yehuda Shalom (2006). "Orbit equivalence rigidity and bounded cohomology". *Ann. of Math. (2)* 164.3, pp. 825–878. MR: 2259246 (cit. on p. 945).

A. Yu. Olshanskii (1980). "An infinite group with subgroups of prime orders". *Izv. Akad. Nauk SSSR Ser. Mat.* 44.2, pp. 309–321, 479. MR: 571100 (cit. on p. 953).

– (1982). "The Novikov-Adyan theorem". *Mat. Sb. (N.S.)* 118(160).2, pp. 203–235, 287. MR: 658789 (cit. on p. 953).

– (1993). "On residualing homomorphisms and G-subgroups of hyperbolic groups". *Internat. J. Algebra Comput.* 3.4, pp. 365–409. MR: 1250244 (cit. on p. 953).

Denis Osin (2006). "Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems". *Mem. Amer. Math. Soc.* 179.843, pp. vi+100. MR: 2182268 (cit. on p. 953).

– (2007). "Peripheral fillings of relatively hyperbolic groups". *Invent. Math.* 167.2, pp. 295–326. MR: 2270456 (cit. on pp. 938, 950).

– (2010). "Small cancellations over relatively hyperbolic groups and embedding theorems". *Ann. of Math. (2)* 172.1, pp. 1–39. MR: 2680416 (cit. on pp. 938, 953, 954).

– (2015). "On acylindrical hyperbolicity of groups with positive first ℓ^2-Betti number". *Bull. Lond. Math. Soc.* 47.5, pp. 725–730. MR: 3403956 (cit. on p. 944).

– (2016). "Acylindrically hyperbolic groups". *Trans. Amer. Math. Soc.* 368.2, pp. 851–888. MR: 3430352 (cit. on pp. 937, 941, 942, 945, 947).

Z. Sela (1997). "Acylindrical accessibility for groups". *Invent. Math.* 129.3, pp. 527–565. MR: 1465334 (cit. on p. 941).

Alessandro Sisto (Dec. 2011). "Contracting elements and random walks". arXiv: 1112.2666 (cit. on p. 943).

Bin Sun (July 2017). "A dynamical characterization of acylindrically hyperbolic groups". arXiv: 1707.04587 (cit. on p. 944).

William P. Thurston (1982). "Three-dimensional manifolds, Kleinian groups and hyperbolic geometry". *Bull. Amer. Math. Soc. (N.S.)* 6.3, pp. 357–381. MR: 648524 (cit. on p. 948).

Kim Whittlesey (2000). "Normal all pseudo-Anosov subgroups of mapping class groups". *Geom. Topol.* 4, pp. 293–307. MR: 1786168 (cit. on p. 952).

Received 2017-12-16.

Denis V. Osin
Department of Mathematics
Vanderbilt University
Nashville 37240
USA
denis.v.osin@vanderbilt.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (959–986)

GLOBAL SURFACES OF SECTION FOR REEB FLOWS IN DIMENSION THREE AND BEYOND

Pedro A. S. Salomão and Umberto L. Hryniewicz

Dedicated to the memory of Professor Kris Wysocki

Abstract

We survey some recent developments in the quest for global surfaces of section for Reeb flows in dimension three using methods from Symplectic Topology. We focus on applications to geometry, including existence of closed geodesics and sharp systolic inequalities. Applications to topology and celestial mechanics are also presented.

1 Introduction

The idea of a global surface of section goes back to Poincaré and the planar circular restricted three-body problem.

Definition 1.1. *Let ϕ^t be a smooth flow on a smooth closed 3-manifold M. An embedded surface $\Sigma \hookrightarrow M$ is a global surface of section for ϕ^t if:*

(i) Each component of $\partial\Sigma$ is a periodic orbit of ϕ^t.

(ii) ϕ^t is transverse to $\Sigma \setminus \partial\Sigma$.

(iii) For every $p \in M \setminus \partial\Sigma$ there exist $t_+ > 0$ and $t_- < 0$ such that $\phi^{t_+}(p)$ and $\phi^{t_-}(p)$ belong to $\Sigma \setminus \partial\Sigma$.

Every $p \in \Sigma \setminus \partial\Sigma$ has a first return time $\tau(p) = \inf\{t > 0 \mid \phi^t(p) \in \Sigma\}$ and the dynamics of the flow are encoded in the first return map

$$\psi : \Sigma \setminus \partial\Sigma \to \Sigma \setminus \partial\Sigma, \qquad \psi(p) = \phi^{\tau(p)}(p). \tag{1}$$

In Poincaré [1912] Poincaré described annulus-like global surfaces of section for the planar circular restricted three-body problem (PCR3BP) for certain values of the Jacobi

MSC2010: primary 37JXX; secondary 53DXX.

constant and mass ratio. Poincaré's global sections motivated his celebrated *last geometric theorem*. The associated first return map preserves an area form, extends up to boundary, and satisfies a twist condition in the range of parameters considered. The exciting discovery made by Poincaré was that the twist condition implies the existence of infinitely many periodic points, i.e., infinitely many periodic orbits for the PCR3BP. In one stroke Poincaré gave a strong push towards a qualitative point of view for studying differential equations, and stated a fixed point theorem intimately connected to the Arnold conjectures and the foundations of Floer Theory.

The recent success of Floer theory and other methods from Symplectic Geometry prompted Hofer to coin the term *Symplectic Dynamics* Bramham and Hofer [2012]. In this note we are concerned with the success of these methods to study Reeb flows in dimension three, with an eye towards applications to geometry.

Our first goal is to discuss existence results for global sections. This will be done in Section 2. After stating Birkhoff's theorem, we focus on Hofer's theory of pseudo-holomorphic curves Hofer [1993]. We survey some published and also some unpublished results, without giving proofs.

Section 3 is devoted to some applications to systolic geometry that were obtained in collaboration with Abbondandolo and Bramham. We will explain how global surfaces of section open the door for symplectic methods in the study of sharp systolic inequalities. We focus on Riemannian two-spheres and on a special case of a conjecture of Viterbo. In Section 4 we present the planar circular restricted three-body problem in more detail. A conjecture due to Birkhoff on the existence of disk-like global surfaces of section for retrograde orbits is discussed.

We intend to convince the reader that there are many positive results for global sections in large classes of flows. However, there are situations where it might be hard to decide whether they exist or not. In sections 5 and 6 we discuss results designed to handle some of these situations. In Section 5 we present deep results of Hofer, Wysocki, and Zehnder [2003] concerning the existence of transverse foliations, and its use in the study of Hamiltonian dynamics near critical levels. In Section 6 we present a Poincaré-Birkhoff theorem for tight Reeb flows on S^3 proved in Hryniewicz, Momin, and Salomão [2015]. It concerns Reeb flows with a pair of closed orbits exactly as those in the boundary of Poincaré's annulus, i.e. forming a Hopf link.

The appendix A discusses a new proof of the existence of infinitely many closed geodesics on any Riemannian two-sphere, which is alternative to the classical arguments of Bangert [1993] and Franks [1992]. It relies on the work of Hingston [1993].

2 Existence results for global surfaces of section

Poincaré constructed his annulus map for a specific family of systems close to integrable[1]. One of the first statements for a large family of systems which can be quite far from integrable is due to Birkhoff.

Theorem 2.1 (Birkhoff [1966]). *Let γ be a simple closed geodesic of a positively curved Riemannian two-sphere. Consider the set A_γ of unit vectors along γ pointing towards one of the hemispheres determined by γ. Then A_γ is a global surface of section for the geodesic flow.*

In other words, every geodesic ray not contained in γ visits both hemispheres infinitely often. We call the embedded annulus A_γ the *Birkhoff annulus*. The family of geodesic flows on positively curved two-spheres is large, making the above statement quite useful. The proof heavily relies on Riemannian geometry and sheds little light on the general existence problem.

A very general theory to attack the existence problem of global surfaces of section exists, and nowadays goes by the name of Schwartzman-Fried-Sullivan theory, see Ghys [2009] or the original works Fried [1982], Schwartzman [1957], and Sullivan [1976]. It produces beautiful theorems with strong conclusions for general flows in dimension three, or even in higher dimensions. The drawback is that these conclusions often require hypotheses which are hard to check, limiting the range of applications. This should not be a surprise because the set of all flows on a 3-manifold is just too wild.

Hofer's pseudo-holomoprhic curve theory deals with the more restrictive class of Reeb flows. However, the results obtained require more reasonable hypotheses which one can often check, as we intend to demonstrate in the next paragraphs. Sometimes results apply automatically for classes of Reeb flows that are large enough to provide applications in topology and geometry. Consider $\mathbb{R}^4$ with coordinates (x_1, y_1, x_2, y_2) and its standard symplectic form $\omega_0 = \sum_{j=1}^{2} dx_j \wedge dy_j$. Here are two examples of such *unconditional* theorems.

Theorem 2.2 (Hofer, Wysocki, and Zehnder [1998]). *The Hamiltonian flow on a smooth, compact and strictly convex energy level in $(\mathbb{R}^4, \omega_0)$ admits a disk-like global surface of section.*

We see Theorem 2.2 as one of the pinnacles of Symplectic Dynamics, it is the guiding application of this theory to the study of global surfaces of section. All results to be discussed in this section are proved using the methods from Hofer, Wysocki, and Zehnder [ibid.].

[1] Angular momentum is preserved in the rotating Kepler problem.

Theorem 2.3 (Hryniewicz [2012, 2014]). *A periodic orbit of the Hamiltonian flow on a smooth, compact and strictly convex energy level in* $(\mathbb{R}^4, \omega_0)$ *bounds a disk-like global surface of section if, and only if, it is unknotted and has self-linking number* -1.

To explain the connection between the above statements and Reeb flows, and to describe further results of this theory, we need first to review basic notions. A contact form λ on a 3-manifold M is a 1-form such that $\lambda \wedge d\lambda$ defines a volume form. Its Reeb vector field R_λ is implicitly defined by

$$d\lambda(R_\lambda, \cdot) = 0, \qquad \lambda(R_\lambda) = 1. \tag{2}$$

The distribution $\xi = \ker \lambda$ is a contact structure, the pair (M, ξ) is a contact manifold. More precisely, these are the co-orientable contact manifolds since λ orients TM/ξ. We only work here with co-orientable contact structures. By a Reeb flow on (M, ξ) we mean one associated to a contact form λ on M such that $\xi = \ker \lambda$. Contact manifolds are the main objects of study in contact topology. Our interest here is shifted towards dynamics.

A knot is called transverse if at every point its tangent space is transverse to the contact structure. A transverse knot with a Seifert surface has a self-linking number, which is invariant under transverse isotopies. It is defined as follows: choose a non-vanishing section of the contact structure along the Seifert surface, then use this section to push the knot off from itself, and finally count intersections with the Seifert surface. The vector bundle $(\xi, d\lambda)$ is symplectic and has a first Chern class $c_1(\xi) \in H^2(M; \mathbb{Z})$. If $c_1(\xi)$ vanishes on $H_2(M; \mathbb{Z})$ then the self-linking number does not depend on the Seifert surface. The book Geiges [2008] by Geiges is a nice reference for these concepts.

Finally, we describe the Conley-Zehnder index in low-dimensions following Hofer, Wysocki, and Zehnder [2003]. Let γ be a periodic trajectory of the flow ϕ^t of the Reeb vector field R_λ, and let $T > 0$ be a period of γ. Since $(\phi^t)^*\lambda = \lambda$, we get a path of $d\lambda$-symplectic linear maps $d\phi^t : \xi_{\gamma(0)} \to \xi_{\gamma(t)}$. The orbit γ is called degenerate in period T if 1 is an eigenvalue of $d\phi^T : \xi_{\gamma(0)} \to \xi_{\gamma(0)}$, otherwise it is called non-degenerate in period T. The contact form λ is called non-degenerate when every periodic trajectory is non-degenerate in every period. When T is the primitive period we may simply call γ degenerate or non-degenerate accordingly.

Since T is a period, we get a well-defined map $\gamma : \mathbb{R}/T\mathbb{Z} \to M$ still denoted by γ without fear of ambiguity. Choose a symplectic trivialization Φ of $\gamma^*\xi$. Then the linearized flow $d\phi^t : \xi_{\gamma(0)} \to \xi_{\gamma(t)}$ gets represented as a path of symplectic matrices $M : \mathbb{R} \to Sp(2)$ satisfying $M(0) = I$, $M(t+T) = M(t)M(T)\ \forall t$. For every non-zero $u \in \mathbb{R}^2$ we write $M(t)u = (r(t)\cos\theta(t), r(t)\sin\theta(t))$ in polar coordinates, for some continuous lift of argument $\theta : \mathbb{R} \to \mathbb{R}$, and define the rotation function $\Delta_M : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ by $\Delta_M(u) = \frac{\theta(T)-\theta(0)}{2\pi}$. The image of Δ_M is a compact interval of length strictly less than $1/2$. The rotation interval J_M is defined as the image of Δ_M.

Consider the following function $\tilde{\mu}(J)$ defined on closed intervals J of length less than $1/2$. If $\partial J \cap \mathbb{Z} = \varnothing$ then set $\tilde{\mu}(J) = 2k$ when $k \in J$, or $\tilde{\mu}(J) = 2k+1$ when $J \subset (k, k+1)$. If $\partial J \cap \mathbb{Z} \neq \varnothing$ then set $\tilde{\mu}(J) = \lim_{\epsilon \to 0^+} \tilde{\mu}(J - \epsilon)$. The Conley-Zehnder index can be finally defined as $\mathrm{CZ}^{\Phi}(\gamma, T) = \tilde{\mu}(J_M)$. We omit the period when it is taken to be the primitive period. If $c_1(\xi)$ vanishes on spheres and $\gamma : \mathbb{R}/T\mathbb{Z} \to M$ is contractible then we write $\mathrm{CZ}^{\mathrm{disk}}$ for the index computed with a trivialization that extends to a capping disk.

The Conley-Zehnder index is an extremely important tool. It is related to Fredholm indices of solutions of many of the elliptic equations from Symplectic Topology, in particular to dimensions of moduli spaces of holomorphic curves.

Definition 2.4 (Hofer, Wysocki and Zehnder)**.** *A contact form λ on a 3-manifold M is dynamically convex if $c_1(\ker \lambda)$ vanishes on spheres and contractible periodic Reeb orbits $\gamma : \mathbb{R}/T\mathbb{Z} \to M$ satisfy $\mathrm{CZ}^{\mathrm{disk}}(\gamma, T) \geq 3$.*

The terminology is justified as follows. The standard contact structure ξ_0 on the unit sphere $S^3 \subset \mathbb{R}^4$ is defined as the kernel of $\lambda_0 = \frac{1}{2}\sum_{j=1}^{2} x_j dy_j - y_j dx_j$ restricted to S^3. More generally, λ_0 restricts to a contact form on any smooth, compact hypersurface S in $(\mathbb{R}^4, \omega_0)$ that is (strictly) star-shaped with respect to the origin. The associated Reeb flow reparametrizes the Hamiltonian flow on S for any Hamiltonian realizing S as a regular energy level. Moreover, it is smoothly conjugated to a Reeb flow on (S^3, ξ_0). Conversely, every Reeb flow on (S^3, ξ_0) is smoothly conjugated to the Reeb flow of λ_0 restricted to some S. When S is strictly convex we get dynamical convexity in view of

Theorem 2.5 (Hofer, Wysocki, and Zehnder [1998])**.** *The Hamiltonian flow on a smooth, compact and strictly convex energy level in $(\mathbb{R}^4, \omega_0)$ is smoothly conjugated to a dynamically convex Reeb flow on (S^3, ξ_0).*

A Reeb flow will be called dynamically convex when it is induced by a dynamically convex contact form. The next result and Theorem 2.5 together imply Theorem 2.3.

Theorem 2.6 (Hryniewicz [2012, 2014])**.** *Let γ be a periodic orbit of a dynamically convex Reeb flow on (S^3, ξ_0). Then γ bounds a disk-like global surface of section if, and only if, it is unknotted and has self-linking number -1. Moreover, such an orbit binds an open book decomposition whose pages are disk-like global surfaces of section.*

These statements are powered by a non-trivial input.

Theorem 2.7 (Hofer, Wysocki, and Zehnder [1996b])**.** *Any Reeb flow on (S^3, ξ_0) has an unknotted periodic orbit with self-linking number -1.*

Putting together theorems 2.6 and 2.7 we obtain a more general version of Theorem 2.2.

Theorem 2.8 (Hofer, Wysocki, and Zehnder [1998]). *Any dynamically convex Reeb flow on (S^3, ξ_0) admits a disk-like global surface of section.*

Global sections open the door for tools in two-dimensional dynamics. Here is a strong application in this direction taken from Hofer, Wysocki, and Zehnder [ibid.]. The return map of the disk obtained from Theorem 2.8 preserves an area form with finite total area. Brouwer's translation theorem provides a periodic orbit simply linked to the boundary of the disk. If the fixed point corresponding to this orbit is removed then we end up with a return map on the open annulus. Results of John Franks [1992] complete the proof of the following statement.

Corollary 2.9 (Hofer, Wysocki, and Zehnder [1998]). *Dynamically convex Reeb flows on (S^3, ξ_0) admit either two or infinitely many periodic orbits.*

To push Theorem 2.6 beyond dynamical convexity one needs to introduce linking assumptions with certain periodic orbits. This is aligned to Schwartzman-Fried-Sullivan theory where one makes linking assumptions with invariant measures.

Theorem 2.10 (Hryniewicz, Licata, and Salomão [2015] and Hryniewicz and Salomão [2011]). *A periodic orbit γ of a Reeb flow on (S^3, ξ_0) binds an open book decomposition whose pages are disk-like global surfaces of section if it matches the following conditions:*

(a) γ is unknotted, has self-linking number -1 and satisfies $\mathrm{CZ}^{\mathrm{disk}}(\gamma) \geq 3$.

(b) γ is linked to all periodic orbits $\gamma' : \mathbb{R}/T\mathbb{Z} \to S^3 \backslash \gamma$ such that either $\mathrm{CZ}^{\mathrm{disk}}(\gamma', T) = 2$, or $\mathrm{CZ}^{\mathrm{disk}}(\gamma', T) = 1$ and γ' is degenerate in period T.

Conversely, if γ is non-degenerate (in its primitive period) then these assumptions are necessary for γ to bound a disk-like global surface of section.

After all these results on the 3-sphere we would like to discuss more general Reeb flows. Can we recover and generalize Birkhoff's Theorem 2.1? To make a statement in this direction we need to recall a few concepts.

The notion of fibered link has a contact topological analogue. If λ is a contact form and L is a transverse link then the right notion of fibered is that L binds an open book decomposition satisfying

(i) $d\lambda$ is an area form on each page, and

(ii) the boundary orientation induced on L by the pages oriented by $d\lambda$ coincides with the orientation induced on L by λ.

Such an open book is said to support the contact structure $\xi = \ker \lambda$. We may call them *Giroux open books* because of their fundamental role in the classification of contact structures due to Giroux [2002]. An open book decomposition is said to be *planar* if pages have no genus. A contact structure orients the underlying 3-manifold by $\lambda \wedge d\lambda$, where λ is any defining contact form. A global surface of section will be called *positive* if the orientation induced on it by the flow and the ambient orientation turns out to orient its boundary along the flow.

Theorem 2.11. *Let (M, ξ) be a closed, connected contact 3-manifold. Let the link $L \subset M$ bind a planar Giroux open book decomposition Θ of M. Denote by $f \in H_2(M, L; \mathbb{Z})$ the class of a page of Θ, and by $\gamma_1, \dots, \gamma_n$ the components of L. Let the contact form λ define ξ and realize L as periodic Reeb orbits, and consider the following assertions:*

- *(i) L bounds a positive genus zero global surface of section for the λ-Reeb flow representing the class f.*
- *(ii) L binds a planar Giroux open book whose pages are global surfaces of section for the λ-Reeb flow and represent the class f.*
- *(iii) The following hold:*
  - *(a) $\mathrm{CZ}^{\Theta}(\gamma_k) > 0$ for all k.*
  - *(b) Every periodic λ-Reeb orbit in $M \setminus L$ has non-zero intersection number with f.*

Then (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). Moreover, (i) $\Rightarrow$ (iii) provided a certain C^∞-generic condition holds.

In (iii-a) $\mathrm{CZ}^{\Theta}(\gamma_k)$ is the Conley-Zehnder index of γ_k in its primitive period computed with a trivialization aligned to the normal of a page of Θ. The genericity needed for (i) $\Rightarrow$ (iii) is implied by non-degeneracy of the contact form. Theorem 2.11 is fruit of joint work with Kris Wysocki and will be proved in Hryniewicz, Salomão, and Wysocki [n.d.]. It heavily relies on Siefring's intersection theory Siefring [2011].

As a first test note that Birkhoff's Theorem 2.1 follows as a consequence. Indeed, the unit sphere bundle of S^2 has a contact form induced by pulling back the tautological 1-from on T^*S^2 via Legendre transform. Reeb flow is geodesic flow. A simple closed geodesic lifts to two closed Reeb orbits, which form a link that binds a supporting open book. Pages are annuli that are isotopic to the Birkhoff annulus. Positivity of the curvature and the Gauss-Bonnet theorem imply that (iii-a) and (iii-b) hold. Birkhoff's theorem follows.

Theorem 2.11 has applications to Celestial Mechanics. The following statement is the abstract result needed for these applications. The standard primitive λ_0 of ω_0 is symmetric by the antipodal map. Identifying antipodal points we obtain $\mathbb{R}P^3 = S^3/\{\pm 1\}$. The restriction of λ_0 to S^3 descends to a contact form on $\mathbb{R}P^3$ defining its standard contact structure, still denoted ξ_0. The Hopf link

$$\widetilde{l}_0 = \{(x_1, y_1, x_2, y_2) \in S^3 \mid x_1 = y_1 = 0 \text{ or } x_2 = y_2 = 0\}$$

is antipodal symmetric and descends to a transverse link l_0 on $\mathbb{R}P^3$. Any transverse link in $(\mathbb{R}P^3, \xi_0)$ transversely isotopic to l_0 will be called a Hopf link. Any transverse knot in $(\mathbb{R}P^3, \xi_0)$ transversely isotopic to a component of l_0 will be called a Hopf fiber.

Theorem 2.12 (Hryniewicz and Salomão [2016], Hryniewicz, Salomão, and Wysocki [n.d.]). *Consider an arbitrary dynamically convex Reeb flow on* $(\mathbb{R}P^3, \xi_0)$. *Any periodic orbit which is a Hopf fiber binds an open book decomposition whose pages are rational disk-like global surfaces of section. Any pair of periodic orbits forming a Hopf link binds an open book decomposition whose pages are annulus-like global surfaces of section.*

These techniques have applications to existence of elliptic periodic orbits. A periodic orbit is elliptic if all Floquet multipliers lie in the unit circle.

Theorem 2.13 (Hryniewicz and Salomão [2016]). *Any Reeb flow on* $(\mathbb{R}P^3, \xi_0)$ *which is sufficiently* C^∞*-close to a dynamically convex Reeb flow admits an elliptic periodic orbit. This orbit binds a rational open book decomposition whose pages are disk-like global surfaces of section. Its double cover has Conley-Zehnder index equal to* 3.

When combined with a result of Harris and Paternain [2008] relating pinched flag curvatures to dynamical convexity, Theorem 2.13 refines the main result from Rademacher [2007].

Corollary 2.14. *Consider a Finsler metric on the two-sphere with reversibility* r. *If all flag curvatures lie in* $(r^2/(r+1)^2, 1]$ *then there exists an elliptic closed geodesic. Moreover, its velocity vector defines a periodic orbit of the geodesic flow that bounds a rational disk-like global surface of section. A fixed point of the return map gives a second closed geodesic.*

We end this section with a topological application. We look for characterizations of contact 3-manifolds in terms of Reeb dynamics, motivated by early fundamental results of Hofer, Wysocki, and Zehnder [1995a, 1999a].

Identify $\mathbb{R}^4 \simeq \mathbb{C}^2$ by $(x_1, y_1, x_2, y_2) \simeq (z_1 = x_1 + iy_1, z_2 = x_2 + iy_2)$ and fix relatively prime integers $p \geq q \geq 1$. The action of $\mathbb{Z}/p\mathbb{Z}$ generated by the map $(z_1, z_2) \mapsto (e^{i2\pi/p} z_1, e^{i2\pi q/p} z_2)$ is free on S^3, and the lens space $L(p,q)$ is defined as its orbit space. The 1-form $\lambda_0 = \frac{1}{2}\sum_{j=1}^2 x_j dy_j - y_j dx_j$ is invariant and descends to a contact form on $L(p,q)$. The induced contact structure is called standard, we still denote it by ξ_0 with no fear of ambiguity.

A knot K on a closed 3-manifold M is p-unknotted if there is an immersion $u : \mathbb{D} \to M$ such that $u|_{\mathbb{D}\setminus\partial\mathbb{D}}$ defines a proper embedding $\mathbb{D} \setminus \partial\mathbb{D} \to M \setminus K$, and $u|_{\partial\mathbb{D}}$ defines a p-covering map $\partial\mathbb{D} \to K$. The map u is called a p-disk for K. The Hopf fiber $S^1 \times 0 \subset S^3$ is $\mathbb{Z}/p\mathbb{Z}$ invariant and descends to the simplest example of a p-unknotted knot in $L(p,q)$. The case $p = 2$ has the following geometric meaning: if we identify $L(2,1)$ with the unit tangent bundle of the round two-sphere then the velocity vector of a great circle is 2-unknotted.

In the presence of a contact structure a transverse p-unknotted knot has a rational self linking number. In the examples given above the knots are transverse and their rational self-linking numbers are equal to $-1/p$. These notions play a role in the following dynamical characterization of standard lens spaces.

Theorem 2.15 (Hofer, Wysocki, and Zehnder [1995a, 1999a] and Hryniewicz, Licata, and Salomão [2015]). *Let (M,ξ) be a closed connected contact 3-manifold, and let $p \geq 1$ be an integer. Then (M,ξ) is contactomorphic to some $(L(p,q),\xi_0)$ if, and only if, it carries a dynamically convex Reeb flow with a p-unknotted self-linking number $-1/p$ periodic orbit.*

This is a special case of more general statements where linking assumptions with certain periodic orbits are used. The existence of a p-unknotted self-linking number $-1/p$ periodic orbit implies that $(M,\xi) = (L(p,q),\xi_0)\#(M',\xi')$ for some contact 3-manifold (M',ξ'). Dynamical convexity forces $(M',\xi') = (S^3,\xi_0)$.

Using that $(L(2,1),\xi_0)$ is contactomorphic to the unit sphere bundle of any Finsler metric on S^2 we get a geometric application. Consider the set $\mathfrak{I}$ of immersions $S^1 \to S^2$ with no positive self-tangencies. Two immersions are declared equivalent if they are homotopic through immersions in $\mathfrak{I}$. This defines an equivalence relation $\sim$ and an element of $\mathfrak{I}/\sim$ will be called a *weak flat knot type*. This notion is related to Arnold's J^+-theory of plane curves. Note that a closed geodesic on a Finsler two-sphere has a well-defined weak flat knot type. Let k_8 be the weak flat knot type of a curve with precisely one self-intersection which is transverse. Clearly there are curves representing k_8 with an arbitrarily large number of self-intersections.

Theorem 2.16 (Hryniewicz and Salomão [2013]). *If a Finsler two-sphere with reversibility r has flag curvatures in $(r^2/(r+1)^2, 1]$ then no closed geodesic represents k_8.*

This statement follows from Theorem 2.15. In fact, the pinching of the curvature forces dynamical convexity (Harris and Paternain [2008]), and the velocity vector of a closed geodesic of type k_8 is unknotted with self-linking number -1 in the unit sphere bundle. Since $\mathbb{R}P^3$ is not the 3-sphere we conclude that such a closed geodesic does not exist.

3 Global surfaces of section applied to systolic geometry

Our first goal in this section is to explain how Birkhoff's annulus-like global surfaces of section (Theorem 2.1) allow for the possibility that symplectic and Riemannian methods be combined to get sharp systolic inequalities on the two-sphere. Our second goal is to describe how disk-like global surfaces of section can be used to prove a special case of Viterbo's conjecture Viterbo [2000]. The results described here were obtained in collaboration with Alberto Abbondandolo and Barney Bramham Abbondandolo, Bramham, Hryniewicz, and Salomão [2017a, 2018, 2017b,c].

The 1-systole $\mathrm{sys}_1(X,g)$ of a closed non-simply connected Riemannian manifold (X,g) is defined as the length of the shortest non-contractible loop. Systolic geometry has its origins in the following results.

Theorem 3.1 (Löwner)**.** *The inequality* $(\mathrm{sys}_1)^2/\mathrm{Area} \leq 2/\sqrt{3}$ *holds for every Riemannian metric on the two-torus. Equality is achieved precisely for the flat torus defined by an hexagonal lattice.*

Theorem 3.2 (Pu)**.** *The inequality* $(\mathrm{sys}_1)^2/\mathrm{Area} \leq \pi/2$ *holds for every Riemannian metric on* $\mathbb{R}P^2$*. Equality is achieved precisely for the round geometry.*

Systolic geometry is a huge and active field, it developed quite a lot since the results of Löwner and Pu. We emphasize Gromov's celebrated paper Gromov [1983].

To include simply connected manifolds one considers the length $\ell_{\min}(X,g)$ of the shortest non-constant closed geodesic of a closed Riemannian manifold (X,g). The systolic ratio is defined by

$$\rho_{\mathrm{sys}}(X,g) = \frac{(\ell_{\min}(X,g))^n}{\mathrm{Vol}(X,g)} \qquad (n = \dim X) \tag{3}$$

The systolic ratio of two-spheres is far from being well understood. An important statement is due to Croke.

Theorem 3.3 (Croke [1988])**.** *The function* $g \mapsto \rho_{\mathrm{sys}}(S^2,g)$ *is bounded among all Riemannian metrics on* S^2*.*

In view of Pu's inequality it is tempting to hope that a round two-sphere (S^2,g_0) maximizes the systolic ratio. Its value is $\rho_{\mathrm{sys}}(S^2,g_0) = \pi$. However, the Calabi-Croke sphere shows that the supremum of $\rho_{\mathrm{sys}}(S^2,g)$ is at least $2\sqrt{3} > \pi$. This is a singular metric constructed by glueing two equilateral triangles along their sides to form a "flat" two-sphere. It can be approximated by smooth positively curved metrics with systolic ratio close to $2\sqrt{3}$.

Question 1. *What is the value of* $\sup_{(S^2,g)} \rho_{\rm sys}(S^2,g)$*? Are there restrictions on the kinds of geometry that approximate this supremum?*

It has been conjectured that the answer to Question 1 is $2\sqrt{3}$. In Balacheff [2010] Balacheff shows that the Calabi-Croke sphere can be seen as some kind of local maximum if non-smooth metrics with a certain type of singular behavior are included.

A *Zoll metric* is one such that all geodesic rays are closed and have the same length. It is interesting that all Zoll metrics on S^2 have conjugated geodesic flows, and have systolic ratio equal to π.

It becomes a natural problem that of understanding the geometry of the function $\rho_{\rm sys}$ near (S^2, g_0). This problem was considered by Babenko and studied by Balacheff. In Balacheff [2006] Balacheff shows that (S^2, g_0) can be seen as a critical point of $\rho_{\rm sys}$ and conjectured that it is a local maximum. We will refer to this conjecture as the Babenko-Balacheff conjecture.

Contact geometry is a natural set-up to study systolic inequalities. This point of view was advertised and used by Álvarez Paiva and Balacheff [2014]. Let α be a contact form on a closed manifold M of dimension $2n-1$ oriented by $\alpha \wedge (d\alpha)^{n-1}$. We denote by $T_{\min}(M,\alpha)$ the minimal period among closed orbits of the Reeb flow. Existence of closed orbits is taken for granted. The contact volume of (M,α) is defined as

$$\mathrm{Vol}(M,\alpha) := \int_M \alpha \wedge (d\alpha)^{n-1}$$

and the systolic ratio of (M,α) as

$$\rho_{\rm sys}(M,\alpha) := \frac{T_{\min}(M,\alpha)^n}{\mathrm{Vol}(M,\alpha)}$$

Note that $\rho_{\rm sys}(M,\alpha)$ is invariant under re-scalings of α.

To see the connection to systolic geometry, consider a Riemannian n-manifold (X,g). The pull-back of the tautological form on T^*X by Legendre transform restricts to a contact form α_g on the unit sphere bundle T^1X. Since the Reeb flow of α_g is the geodesic flow of g, we get $T_{\min}(T^1X,\alpha_g) = \ell_{\min}(X,g)$. It turns out that $\mathrm{Vol}(X,g)$ and $\mathrm{Vol}(T^1X,\alpha_g)$ are proportional by a constant depending only on n. Hence $\rho_{\rm sys}(T^1X,\alpha_g) = C_n\, \rho_{\rm sys}(X,g)$ for every Riemannian metric g on X, where C_n depends only on n.

A Zoll contact form is one such that all Reeb trajectories are periodic and have the same period. These are usually called regular in the literature, but we prefer the term Zoll in view of the above connection to the Riemannian case.

A convex body in $\mathbb{R}^{2n}$ is a compact convex set with non-empty interior. In Viterbo [2000] Viterbo conjectured that

$$\frac{c(K)^n}{n!\mathrm{Vol}(K)} \leq 1 \tag{4}$$

holds for every convex body $K \subset \mathbb{R}^{2n}$ and every symplectic capacity c, where $\mathrm{Vol}(K)$ denotes euclidean volume. We end by discussing a special case of the conjecture. Let K be a convex body in $(\mathbb{R}^{2n}, \omega_0)$ with smooth and strictly convex boundary, with the origin in its interior. Denote by $\iota : \partial K \to \mathbb{R}^{2n}$ the inclusion map, and by λ_0 the standard Liouville form $\lambda_0 = \frac{1}{2}\sum_{j=1}^n x_j dy_j - y_j dx_j$. Then $\iota^*\lambda_0$ is a contact form on ∂K. In Hofer and Zehnder [1994] it is claimed that the Hofer-Zehnder capacity of K is equal to $T_{\min}(\partial K, \iota^*\lambda_0)$. In this case (4) is restated as

$$\rho_{\mathrm{sys}}(\partial K, \iota^*\lambda_0) \leq 1 \tag{5}$$

which is supposed to be an equality if, and only if, $\iota^*\lambda_0$ is Zoll.

Having described our problems, we move on to state some results. Recall that for $\delta \in (0, 1]$, a positively curved closed Riemannian manifold is said to be δ-pinched if the minimal and maximal values $K_{\min}, K_{\max}$ of the sectional curvatures satisfy $K_{\min}/K_{\max} \geq \delta$. On a positively curved two-sphere we write $\ell_{\max}$ for the length of the longest closed geodesic without self-intersections. Note that $\ell_{\max}$ is finite.

Theorem 3.4 (Abbondandolo, Bramham, Hryniewicz, and Salomão [2017a])**.** *If (S^2, g) is δ-pinched for some $\delta > (4+\sqrt{7})/8 = 0.8307...$ then $\ell_{\min}(S^2, g)^2 \leq \pi \mathrm{Area}(S^2, g) \leq \ell_{\max}(S^2, g)^2$. Moreover, any of these inequalities is an equality if, and only if, the metric is Zoll.*

This first inequality confirms the Babenko-Balacheff conjecture on an explicit and somewhat large C^2-neighborhood of the round geometry. It seems that the upper bound involving $\ell_{\max}$ was not known before.

We discuss some related problems before explaining the role of global surfaces of section in the proof of Theorem 3.4. The pinching constant δ seems to be a helpful parameter. For instance, one could consider the non-increasing bounded (Theorem 3.3) function

$$\rho : (0, 1] \to \mathbb{R} \qquad \rho(\delta) = \sup\{\rho_{\mathrm{sys}}(S^2, g) \mid (S^2, g) \text{ is } \delta\text{-pinched}\}$$

to study the positively curved case.

Question 2. *Is it true that $\rho(1/4) = \pi$? What does the graph of $\rho(\delta)$ look like?*

The Calabi-Croke sphere shows that $\lim_{\delta \to 0^+} \rho(\delta) \geq 2\sqrt{3}$. Theorem 3.4 implies that $\rho(\delta) = \pi$ for all $\delta > (4+\sqrt{7})/8$. One must try to understand among which metrics does the round metric maximize systolic ratio. Assuming positive curvature it might be reasonable to expect that $\inf\{\delta \mid \rho(\delta) = \pi\} \leq 1/4$.

If curvature assumptions are dropped then the situation might be much harder. What about symmetry assumptions? Here is a result in this direction that answers a question by Álvarez-Paiva and Balacheff.

Theorem 3.5. *Inequality $\rho_{\mathrm{sys}} \leq \pi$ holds for every sphere of revolution, with equality precise when the metric is Zoll.*

Global surfaces of section show up in the proofs of theorems 3.4 and 3.5 to connect systolic inequalities to a quantitative fixed point theorem for symplectic maps of the annulus. We outline the proof to make this point precise.

Let (S^2, g) be δ-pinched. If $\delta > 1/4$ then $\ell_{\min}$ is only realized by simple closed geodesics. Let γ be a closed geodesic of length $\ell_{\min}$. By Theorem 2.1 the Birkhoff annulus A_γ is a global surface of section. Let λ be the 1-form on A_γ given by restricting the contact form α_g. Then $d\lambda$ is an area form on the interior of A_γ, and vanishes on ∂A_γ. The total $d\lambda$-area of A_γ is $2\ell_{\min}$.

The first return map ψ and the first return time τ are defined on the interior of A_γ, but it turns out that they extend smoothly to A_γ. Moreover, ψ preserves boundary components. Santaló's formula reads

(6) $$2\pi \operatorname{Area}(S^2, g) = \int_{T^1S^2} \alpha_g \wedge d\alpha_g = \int_{A_\gamma} \tau \, d\lambda$$

Since ψ preserves the 2-form $d\lambda$, it follows that $\psi^*\lambda - \lambda$ is closed.

We now need to consider lifts of ψ to the universal covering of A_γ. If ψ admits a lift in the kernel of the FLUX homomorphism then $\psi^*\lambda - \lambda$ is exact. The unique primitive σ of $\psi^*\lambda - \lambda$ satisfying

$$\sigma(p) = \int_p^{\psi(p)} \lambda \quad \forall p \in \partial A_\gamma$$

is called the *action of* ψ. Here the integral is taken along the boundary according to the lift with zero FLUX. The Calabi invariant is defined as

$$\operatorname{CAL}(\psi) = \frac{1}{\int_{A_\gamma} d\lambda} \int_{A_\gamma} \sigma \, d\lambda = \frac{1}{2\ell_{\min}} \int_{A_\gamma} \sigma \, d\lambda$$

Of course, we need to worry about whether ψ admits a lift of zero FLUX, but this follows from reversibility of the geodesic flow.

It is a very general fact that τ is also a primitive of $\psi^*\lambda - \lambda$. Toponogov's theorem proves that if $\delta > 1/4$ then

(7) $$\tau = \sigma + \ell_{\min}$$

Combining (7) with (6) we finally get

$$2\pi \,\mathrm{Area}(S^2, g) = 2(\ell_{\min})^2 + 2\ell_{\min}\mathrm{CAL}(\psi) \tag{8}$$

Equations (7) and (8) should be seen as some kind of dictionary between geometry and dynamics: action corresponds to length, Calabi invariant corresponds to area.

We are now in position to make the link to the quantitative fixed point theorem and conclude the argument. Roughly speaking, the theorem states:

If ψ admits a generating function (of a specific kind), $\mathrm{CAL}(\psi) \le 0$ and $\psi \neq id$, then there exists a fixed point p_0 satisfying $\sigma(p_0) < 0$.

Arguing indirectly, suppose that either $\pi\,\mathrm{Area} < (\ell_{\min})^2$, or $\pi\,\mathrm{Area} = (\ell_{\min})^2$ and g is not Zoll. It follows from (8) and a little more work that either $\mathrm{CAL}(\psi) < 0$, or $\mathrm{CAL}(\psi) = 0$ and ψ is not the identity. Toponogov's theorem comes into play again to show that ψ admits the required generating function provided $\delta > (4+\sqrt{7})/8$. The fixed point theorem applies to give a fixed point of negative action. By (7) this fixed point corresponds to a closed geodesic of length strictly smaller than $\ell_{\min}$. This contradiction finishes the proof.

The above argument reveals how global surfaces of section can serve as bridge between systolic geometry and symplectic dynamics. The same strategy proves a special case of Viterbo's conjecture in dimension 4.

Theorem 3.6 (Abbondandolo, Bramham, Hryniewicz, and Salomão [2018]). *There exists a C^3-neighborhood $\mathcal{U}$ of the space of Zoll contact forms on S^3 such that $\alpha \in \mathcal{U} \Rightarrow \rho_{\mathrm{sys}}(S^3, \alpha) \le 1$ with equality if, only if, α is Zoll.*

The proof again strongly relies on global surfaces of sections. Namely, if a contact form is C^3-close to the standard contact form λ_0 then its Reeb flow admits a disk-like global surface of section whose first return map extends up to the boundary and is C^1-close to the identity. We have a dictionary between maps and flows just as in the proof of Theorem 3.4: contact volume corresponds to Calabi invariant, return time corresponds to action. The quantitative fixed point theorem applies to give the desired conclusion.

One could see the constants in sharp systolic inequalities for Riemannian surfaces as invariants. Similarly, one could hope to construct contact invariants from sharp systolic inequalities for contact forms. The following statement shows that this is not possible in dimension three: systolic inequalities are not purely contact topological phenomena. For example, inequalities such as (5) must depend on the convexity assumption.

Theorem 3.7 (Abbondandolo, Bramham, Hryniewicz, and Salomão [2018, 2017b]). *For every co-orientable contact 3-manifold (M, ξ) and every $c > 0$ there exists a contact form α on M satisfying $\xi = \ker\alpha$ and $\rho_{\mathrm{sys}}(M, \alpha) > c$.*

Hofer, Wysocki, and Zehnder [1999a, 1998] introduced the notion of dynamically convex contact forms, see Section 2 for a detailed discussion. It plays a crucial role in the construction of global surfaces of section (theorems 2.6, 2.8). Dynamical convexity is automatically satisfied on the boundary of a smooth convex body with strictly convex boundary. It becomes relevant to decide whether (5) holds for dynamically convex contact forms on S^3.

Theorem 3.8 (Abbondandolo, Bramham, Hryniewicz, and Salomão [2017c])**.** *Given any* $\epsilon > 0$ *there is a dynamically convex contact form* α *on* S^3 *such that* $\rho_{\text{sys}}(S^3, \alpha) > 2 - \epsilon$.

A narrow connection between high systolic ratios and negativity of Conley-Zehnder indices is quantified in Abbondandolo, Bramham, Hryniewicz, and Salomão [ibid.].

Observe that Theorem 3.8 implies that either Viterbo's conjecture is not true, or there exists a dynamically convex contact form on S^3 whose Reeb flow is not conjugated to the Reeb flow on a strictly convex hypersurface of $(\mathbb{R}^4, \omega_0)$. Unfortunately we can not decide which alternative holds. It also proves that there are smooth compact star-shaped domains U in $(\mathbb{R}^4, \omega_0)$ with the following property: the value $c(U)$ of any capacity realized as the action of some closed characteristic on ∂U is strictly larger than the Gromov width of U.

Global surfaces of section continue to play essential role in the proofs of Theorem 3.7 and Theorem 3.8. Both start by constructing global sections for certain Reeb flows with well-controlled return maps. Then the Reeb flows are modified by carefully changing the return maps in order to make the systolic ratio increase.

4 The planar circular restricted three-body problem

The three-body problem is that of understanding the motion of three massive particles which attract each other according to Newton's law of gravitation. Some simplifying assumptions turn this problem into a two-degree-of-freedom Hamiltonian system:

- The three particles move in a fixed plane.
- The mass of the third body (satellite) is neglected and so the first two particles (primaries) move according to the two-body problem.
- The primaries move on circular trajectories about their center of mass.

In inertial coordinates where the center of mass of the primaries rests at the origin one gets $z_1 = r_1 e^{i\omega t}$ and $z_2 = -r_2 e^{i\omega t}$ for some ω, where $r_1, r_2 > 0$ satisfy $m_1 r_1 - m_2 r_2 = 0$ and $(r_1 + r_2)^3 \omega^2 = m_1 + m_2$. It is harmless to assume that $\omega = r_1 + r_2 = m_1 + m_2 = 1$ which makes the mass ratio $\mu := m_1 = r_2 \in (0, 1)$ the unique parameter of the system.

In rotating (non-inertial) coordinates the position $q(t) \in \mathbb{C}$ of the satellite relative to the second primary is given by $z_3(t) = (q(t) - \mu)e^{it}$, from where it follows that

$$\ddot{q} + 2i\dot{q} - (q - \mu) = -\mu \frac{q-1}{|q-1|^3} - (1-\mu)\frac{q}{|q|^3}. \tag{9}$$

As is well known, if we set $p = \dot{q} + i(q - \mu)$ and consider

$$H_\mu(q,p) = \frac{1}{2}|p|^2 + \langle q - \mu, ip\rangle - \frac{\mu}{|q-1|} - \frac{1-\mu}{|q|}, \tag{10}$$

then (9) becomes Hamilton's equations

$$\dot{q} = \nabla_p H_\mu, \qquad \dot{p} = -\nabla_q H_\mu. \tag{11}$$

The function H_μ has five critical points. A sublevel set below its lowest critical value defines three *Hill regions* in the configuration plane, two of which are bounded while the third is a neighborhood of ∞. Each bounded Hill region is topologically a punctured disk and contains a primary, namely, one of them is a punctured neighborhood of the origin and the other is a punctured neighborhood of 1. The boundaries of the Hill regions are called *ovals of zero velocity*, since there we have $(q - \mu) = -ip \Leftrightarrow \dot{q} = 0$. From now on we restrict to *subcritical cases*, i.e. energy levels $H_\mu = -c$ where $-c$ is below the lowest critical value of H_μ. We focus on the bounded Hill region near the origin.

Following Poincaré, mathematicians first tried to understand the limiting behavior as $\mu \to 0^+$ or as $\mu \to 1^-$. The limit as $\mu \to 0^+$ is in some ways better behaved than the limit $\mu \to 1^-$, but sometimes it is just the other way around. In the limit $\mu = 0$ the system describes the so-called rotating Kepler problem, where all mass is concentrated at the origin. The boundary of the bounded Hill region about the origin converges to a circle of definite radius. As $\mu \to 1^-$ the bounded Hill region about the origin collapses, and we face a somewhat more singular situation.

Definition 4.1. *A retrograde orbit is a periodic orbit $t \mapsto (q(t), p(t))$ such that $q(t)$ is in the Hill region about the origin, and describes a curve without self-intersections with winding number -1 around the origin. Analogously, a direct orbit is a periodic orbit $t \mapsto (q(t), p(t))$ such that $q(t)$ is in the Hill region about the origin, and describes a curve without self-intersections with winding number $+1$ around the origin.*

The difficulty in finding direct orbits led Birkhoff to consider the following strategy in Birkhoff [1914, section 19]. Firstly one should try to find a disk-like global surface of section bounded by a (doubly covered) retrograde orbit. For this to make sense collision orbits need to be regularized. Secondly, due to preservation of an area form with finite total area, one can apply Brouwer's translation theorem to the first return map and find a fixed point that should correspond to a direct orbit. Two main difficulties are: (1) for an arbitrary mass ratio it is hard to find global surfaces of section, and (2) it might be hard to check that the fixed point corresponds to a direct orbit. The following is extracted from Birkhoff [ibid., section 19]:

"This state of affairs seems to me to make it probable that the restricted problem of three bodies admit of reduction to the transformation of a discoid into itself as long as there is a closed oval of zero velocity about J, and that in consequence there always exists at least one direct periodic orbit of simple type."

More recently this has been called a conjecture, which perhaps should be read as following: For any value of μ and any subcritical energy value, there must be a way of finding a disk-like global surface of section in order to understand the movement of the satellite inside the Hill region about the origin. To implement the strategy of Birkhoff this disk should be spanned by the retrograde orbit, in particular fixed points could be good candidates for direct orbits. Again, all this only makes sense if collision orbits are regularized.

Note that the smallest critical value of H_μ converges to $-\frac{3}{2}$ both when $\mu \to 0^+$ or $\mu \to 1^-$. Here is a good point to state and discuss our result concerning Birkhoff's conjecture.

Theorem 4.2. *For every $c > \frac{3}{2}$ there exists $\epsilon > 0$ such that the following holds.*

(a) If $1 - \mu < \epsilon$ and collisions are regularized via Levi-Civita regularization, then the double cover of every retrograde orbit inside the Hill region about the origin bounds a disk-like global surface of section. Moreover, if we quotient by antipodal symmetry then this disk descends to a rational disk-like global surface of section.

(b) If $\mu < \epsilon$ and collisions are regularized via Moser regularization, then every retrograde orbit inside the Hill region about the origin bounds a rational disk-like global surface of section.

Results of Albers, Fish, Frauenfelder, Hofer and van Koert from Albers, Fish, Frauenfelder, Hofer, and van Koert [2012] imply that if $1 - \mu$ is small enough then Levi-Civita regularization lifts the Hamiltonian flow on the corresponding component of $H_\mu^{-1}(-c)$ to the characteristic flow on a strictly convex hypersurface $\widetilde{\Sigma}_{\mu,c}$, up to time reparametrization. Moreover, $\widetilde{\Sigma}_{\mu,c}$ is antipodal symmetric and each state is represented twice as a pair of antipodal points. Results from Hofer, Wysocki, and Zehnder [1998] apply and give disk-like global surfaces of section in $\widetilde{\Sigma}_{\mu,c}$. Statement (a) above says that there is such a global section in $\widetilde{\Sigma}_{\mu,c}$ spanned by the lift of every doubly covered retrograde orbit, and that it descends to a global section in the quotient $\Sigma_{\mu,c} = \widetilde{\Sigma}_{\mu,c}/\{\pm 1\}$. If $\mu = 0$ then Moser regularization applies to the rotating Kepler problem to compactify the Hamiltonian flow on $H_\mu^{-1}(-c)$ to the characteristic flow on a fiberwise starshaped hypersurface $\Sigma_{\mu,c}$ inside TS^2, up to time reparametrization. A proof of this statement can be found in the paper Albers, Frauenfelder, van Koert, and Paternain [2012] where the contact-type property of energy levels of the PCR3BP is studied. Again we have $\Sigma_{\mu,c} \simeq \mathbb{R}P^3$. Statement (b) above says that every retrograde orbit bounds a rational disk-like global surface of section in $\Sigma_{\mu,c}$. A proof in this case would rely on the dynamical convexity obtained in Albers, Fish, Frauenfelder, and van Koert [2013] for $\mu = 0$. Hence, for μ close to 0 or 1 we can always apply Theorem 2.13 and obtain a pair of periodic orbits which are 2-unknotted and have self-linking number $-1/2$. These orbits are transversely isotopic to (a quotient of) a Hopf link. Theorem 2.12 can also be applied and an annulus-like global surface of section is obtained.

We end with a sketch of proof of (a) in Theorem 4.2. Fix $c > 3/2$. The component $\dot{\Sigma}_{\mu,c} \subset H_\mu^{-1}(-c)$ which projects to the Hill region surrounding $0 \in \mathbb{C}$ contains collision orbits. These orbits are regularized with the aid of Levi-Civita coordinates $(v,u) \in \mathbb{C}\times\mathbb{C}$ given by $q = 2v^2$ and $p = -\frac{u}{\bar{v}}$, which are symplectic up to a constant factor. The regularized Hamiltonian is

$$\begin{aligned} K_{\mu,c}(v,u) &:= |v|^2(H_\mu(p,q)+c) \\ &= \frac{1}{2}|u|^2 + 2|v|^2\langle u, iv\rangle - \mu\Im(uv) - \frac{1-\mu}{2} - \mu\frac{|v|^2}{|2v^2-1|} + c|v|^2, \end{aligned} \tag{12}$$

and there is a two-to-one correspondence between a centrally symmetric sphere-like component $\widetilde{\Sigma}_{\mu,c} \subset K_{\mu,c}^{-1}(0)$ and $\dot{\Sigma}_{\mu,c}$, up to collisions.

Now we consider the re-scaled coordinates $v = \hat{v}\sqrt{1-\mu}$ and $u = \hat{u}\sqrt{1-\mu}$, with Hamiltonian

$$\begin{aligned} \hat{K}_{\mu,c}(\hat{v},\hat{u}) &:= \frac{1}{1-\mu}K_{\mu,c}(v,u) \\ &= \frac{1}{2}|\hat{u}|^2 + 2(1-\mu)|\hat{v}|^2\langle \hat{u}, i\hat{v}\rangle - \mu\Im(\hat{u}\hat{v}) - \frac{1}{2} - \mu\frac{|\hat{v}|^2}{|2(1-\mu)\hat{v}^2-1|} + c|\hat{v}|^2. \end{aligned} \tag{13}$$

The component $\widetilde{\Sigma}_{\mu,c} \subset K_{\mu,c}^{-1}(0)$ gets re-scaled and we denote it by $\hat{\Sigma}_{\mu,c} \subset \hat{K}_{\mu,c}^{-1}(0)$.

Taking $\mu \to 1^-$ we see from (13) that $\hat{\Sigma}_{\mu,c}$ converges in the C^∞ topology to a hypersurface satisfying

$$\frac{1}{2}|\hat{u}|^2 - \Im(\hat{u}\hat{v}) + (c-1)|\hat{v}|^2 = \frac{1}{2}. \tag{14}$$

In order to have a better picture of the hypersurface in (14), we denote, for simplicity, $\hat{v} = \hat{v}_1 + i\,\hat{v}_2$ and $\hat{u} = \hat{u}_1 + i\,\hat{u}_2$. Then (14) is equivalent to

$$(\hat{u}_1 - \hat{v}_2)^2 + (\hat{u}_2 - \hat{v}_1)^2 + 2\left(c - \frac{3}{2}\right)(\hat{v}_1^2 + \hat{v}_2^2) = 1. \tag{15}$$

Taking new coordinates $(w = w_1 + i w_2, z = z_1 + i z_2) \in \mathbb{C} \times \mathbb{C}$ with $w_1 = \hat{u}_1 - \hat{v}_2$, $w_2 = \hat{u}_2 - \hat{v}_1$ and $z = \hat{v}\sqrt{2c-3}$, which are symplectic up to a constant factor, we see that (15) is equivalent to $w_1^2 + w_2^2 + z_1^2 + z_2^2 = 1$.

We conclude that the regularized Hamiltonian flow on $\hat{\Sigma}_{\mu,c}$ converges smoothly to the standard Reeb flow on (S^3, ξ_0) as $\mu \to 1^-$ up to reparametrizations. Its orbits are the Hopf fibers. Since the projection of the retrograde orbit winds once around $0 \in \mathbb{C}$ in q-coordinates, it is doubly covered by a simple closed orbit $P_{\mu,c} \subset \widetilde{\Sigma}_{\mu,c}$, which in z-coordinates winds once around $0 \in \mathbb{C}$. Hence, $P_{\mu,c}$ converges smoothly to a Hopf fiber in (w, z) and, in particular, it is unknotted and has self-linking number -1. The dynamical convexity of the Hamiltonian flow on $\hat{\Sigma}_{\mu,c}$ and Theorem 2.6 imply that it is the boundary of a disk-like global surface of section. In view of Theorem 2.12, we may assume that this global section descends to a rational disk-like global section on $\Sigma_{\mu,c} = \widetilde{\Sigma}_{\mu,c}/\{\pm 1\}$.

5 Transverse foliations

We discuss the idea of transverse foliations adapted to a 3-dimensional flow based on the concepts introduced by Hofer, Wysocki and Zehnder in Hofer, Wysocki, and Zehnder [2003]. This generalizes the notion of open books and global sections.

Definition 5.1. *Let ϕ^t be a smooth flow on an oriented closed 3-manifold M. A transverse foliation for ϕ^t is formed by:*

(i) A finite set $\mathcal{P}$ of primitive periodic orbits of ϕ^t, called binding orbits.

(ii) A smooth foliation of $M \setminus \cup_{P \in \mathcal{P}} P$ by properly embedded surfaces. Every leaf $\dot{\Sigma}$ is transverse to ϕ^t, has an orientation induced by ϕ^t and M, and there exists a compact embedded surface $\Sigma \hookrightarrow M$ so that $\dot{\Sigma} = \Sigma \setminus \partial\Sigma$ and $\partial\Sigma$ is a union of components of $\cup_{P \in \mathcal{P}} P$. An end z of $\dot{\Sigma}$ is called a puncture. To each puncture z there is an associated component $P_z \in \mathcal{P}$ of $\partial\Sigma$ called the asymptotic limit of $\dot{\Sigma}$ at z. A puncture z of $\dot{\Sigma}$ is called positive if the orientation on P_z induced by $\dot{\Sigma}$ coincides with the orientation induced by ϕ^t. Otherwise z is called negative.

The following theorem is a seminal result on the existence of transverse foliations for Reeb flows on the tight 3-sphere. It is based on pseudo-holomorphic curve theory in symplectic cobordisms.

Theorem 5.2 (Hofer, Wysocki, and Zehnder [2003]). *Let ϕ^t be a nondegenerate Reeb flow on (S^3, ξ_0). Then ϕ^t admits a transverse foliation. The binding orbits have self-linking number -1 and their Conley–Zehnder indices are 1, 2 or 3. Every leaf $\dot\Sigma$ is a punctured sphere and has precisely one positive puncture. One of the following conditions holds:*

- *The asymptotic limit of $\dot\Sigma$ at its positive puncture has Conley-Zehnder index 3 and the asymptotic limit of $\dot\Sigma$ at any negative puncture has Conley-Zehnder index 1 or 2. There exists at most one negative puncture whose asymptotic limit has Conley-Zehnder index 2.*
- *The asymptotic limit of $\dot\Sigma$ at its positive puncture has Conley-Zehnder index 2 and the asymptotic limit of $\dot\Sigma$ at any negative puncture has Conley-Zehnder index 1.*

The open books with disk-like pages constructed in Hofer, Wysocki, and Zehnder [1995a, 1999a, 1998], Hryniewicz [2012, 2014], Hryniewicz, Licata, and Salomão [2015], and Hryniewicz and Salomão [2011] for Reeb flows on (S^3, ξ_0) are particular cases of transverse foliations with a single binding orbit. The main obstruction for the existence of such an open book with a prescribed binding orbit P is the presence of closed orbits with Conley-Zehnder index 2 which are unlinked to P. One particular transverse foliation of interest which deals with such situations is the so called 3-2-3 foliation.

Definition 5.3. *A 3-2-3 foliation for a Reeb flow ϕ^t on (S^3, ξ_0) is a transverse foliation for ϕ^t with precisely three binding orbits P_3, P_2 and P_3'. They are unknotted, mutually unlinked and their respective Conley-Zehnder indices are 3, 2 and 3. The leaves are punctured spheres and consist of*

- *A pair of planes U_1 and U_2, both asymptotic to P_2 at their positive punctures.*
- *A cylinder V asymptotic to P_3 at its positive puncture and to P_2 at its negative puncture; a cylinder V' asymptotic to P_3' at its positive puncture and to P_2 at its negative puncture.*
- *A one parameter family of planes asymptotic to P_3 at their positive punctures; a one parameter family of planes asymptotic to P_3' at their positive punctures.*

The 3-2-3 foliations are the natural objects to consider if one studies Hamiltonian dynamics near certain critical energy levels.

Take a Hamiltonian H on $\mathbb{R}^4$ which has a critical point $p \in H^{-1}(0)$ with Morse index 1 and of saddle-center type. Its center manifold is foliated by the so called Lyapunoff orbits $P_{2,E} \subset H^{-1}(E)$, $E > 0$ small. Each one of them is unknotted, hyperbolic inside its energy level and has Conley-Zehnder index 2.

Assume that for every $E < 0$ the energy level $H^{-1}(E)$ contains two sphere-like components S_E and S_E' which develop a common singularity at p as $E \to 0^-$. This means that S_E converges in the Hausdorff topology to $S_0 \subset H^{-1}(0)$ as $E \to 0^-$, where S_0 is homeomorphic to the 3-sphere and contains p as its unique singularity. The analog holds for S'_E. Therefore, $S_0 \cap S'_0 = \{p\}$ and, for $E > 0$ small, $H^{-1}(E)$ contains a sphere-like component W_E close to $S_0 \cup S'_0$. We observe that W_E contains the Lyapunoff orbit $P_{2,E}$ and is in correspondence with the connected sum of S_E and S'_E.

Definition 5.4. *We say that S_0 is strictly convex if S_0 bounds a convex domain in $\mathbb{R}^4$ and all the sectional curvatures of $S_0 \setminus \{p\}$ are positive. We say that S'_0 is strictly convex if analogous conditions hold.*

The following theorem is inspired by results in Hofer, Wysocki, and Zehnder [2003].

Theorem 5.5 (de Paulo and Salomão [2018, n.d.])**.** *If H is real analytic and both S_0 and S'_0 are strictly convex then, for every $E > 0$ small, the Hamiltonian flow on the sphere-like component $W_E \subset H^{-1}(E)$ admits a 3-2-3 foliation. The Lyapunoff orbit $P_{2,E}$ is one of the binding orbits and there exist infinitely many periodic orbits and infinitely many homoclinics to $P_{2,E}$ in W_E.*

One difficulty in proving Theorem 5.5 is that there are no non-degeneracy assumptions of any kind. A criterium for checking strict convexity of the subsets S_0 and S'_0 is found in Salomão [2003].

The notion of 3-2-3 foliation is naturally extended to Reeb flows on connected sums $\mathbb{R}P^3\#\mathbb{R}P^3$. In this case the binding orbits P_3 and P'_3 are non-contractible and the families of planes are asymptotic to their respective double covers. The existence of 3-2-3 foliations for Reeb flows on $\mathbb{R}P^3\#\mathbb{R}P^3$ is still an object of study and it is conjectured that they exist for some Hamiltonians in celestial mechanics such as the Euler's problem of two centers in the plane and the planar circular restricted three body problem for energies slightly above the first Lagrange value.

A more general theory of transverse foliations for Reeb flows still needs to be developed. If one wishes to use holomorphic curves then one step is implemented by Fish and Siefring [2013], who showed persistence under connected sums. Transverse foliations on mapping tori of disk-maps were constructed by Bramham [2015a,b] to study questions about rigidity of pseudo-rotations.

6 A Poincaré-Birkhoff theorem for tight Reeb flows on S^3

Poincaré's last geometric theorem is nowadays known as the Poincaré-Birkhoff theorem. In its simplest form it is a statement about fixed points of area-preserving annulus homeomorphisms $f : \mathbb{R}/\mathbb{Z} \times [0, 1] \to \mathbb{R}/\mathbb{Z} \times [0, 1]$ preserving orientation and boundary components. The map f can be lifted to the universal covering $\mathbb{R} \times [0, 1]$. Let us denote projection onto the first coordinate by $p : \mathbb{R} \times [0,1] \to \mathbb{R}$. Then f is said to satisfy a *twist condition* on the boundary if it admits a lift to the universal covering $F : \mathbb{R} \times [0,1] \to \mathbb{R} \times [0,1]$ such that the rotation numbers

$$\lim_{n\to\infty} \frac{p \circ F^n(x,0)}{n} \qquad \lim_{n\to\infty} \frac{p \circ F^n(x,1)}{n}$$

differ. We call the open interval I bounded by these numbers the twist interval.

Theorem 6.1 (Poincaré-Birkhoff Birkhoff [1913] and Poincaré [1912]). *If $I \cap \mathbb{Z} \neq \varnothing$ then f has at least two fixed points.*

Poincaré [1912] found annulus-like global surfaces of section for the PCR3BP for energies below the lowest critical value of the Hamiltonian, and when the mass is almost all concentrated in the primary around which the satellite moves. The boundary orbits form a Hopf link in the three-sphere. For generic values of the parameters, the Poincaré-Birkhoff theorem applies to the associated return map and proves the existence of infinitely many periodic orbits.

One also finds such pair of orbits for the Hamiltonian flow on a smooth, compact and strictly convex energy level inside $(\mathbb{R}^4, \omega_0)$. In fact, the fundamental result of Hofer, Wysocki, and Zehnder [1998] provides an unknotted periodic orbit P_0 that bounds a disk-like global surface of section. Brouwer's translation theorem yields a second periodic orbit P_1 simply linked to P_0, but much more can be said. The orbit P_0 is the binding of an open book decomposition whose pages are disk-like global surfaces of section. It turns out that the following statement follows: the flow is smoothly conjugated to a Reeb flow on (S^3, ξ_0) in such a way that $P_0 \cup P_1$ corresponds to a link transversely isotopic to the standard Hopf link

$$\widetilde{l}_0 = \{(x_1, y_1, x_2, y_2) \in S^3 \mid x_1 = y_1 = 0 \text{ or } x_2 = y_2 = 0\}$$

If the fixed point corresponding to P_1 is removed from the disk-like global section spanned by P_0, then we obtain a diffeomorphism of the open annulus that preserves a standard area-form and can be continuously extended to the boundary. It is interesting to study the twist condition for this map. We need to consider the transverse rotation numbers θ_0 and θ_1 of P_0 and P_1 with respect to Seifert surfaces (disks). In terms of Conley-Zehnder indices, these can be read as follows:

$$1 + \theta_0 = \lim_{n\to\infty} \frac{\mathrm{CZ}(P_0^n)}{2n} \qquad 1 + \theta_1 = \lim_{n\to\infty} \frac{\mathrm{CZ}(P_1^n)}{2n} \tag{16}$$

Here $\mathrm{CZ}(P_i^n)$ denotes the Conley-Zehnder index of the n-iterated orbit P_i computed with respect to a global trivialization of ξ_0. The open book singles out a lift of the map to the strip such that the rotation numbers on the boundary are precisely $1/\theta_0$ and θ_1. The Poincaré-Birkhoff theorem proves the following non-trivial statement: *If $\theta_1 \neq 1/\theta_0$ then there are infinitely many periodic orbits in the complement of $P_0 \cup P_1$. These orbits are distinguished by their homotopy classes in the complement of $P_0 \cup P_1$.*

One motivation for the main result of this section is to study the possibility of extending the above discussion to situations where neither P_0 nor P_1 bound global sections. Before the statement we need some notation. The term *Hopf link* will be referred to any transverse link in (S^3, ξ_0) that is transversely isotopic to the standard Hopf link $\widetilde{l}_0$. Given non-zero vectors $u, v \in \mathbb{R}^2$ in the complement of the third quadrant, we write $u > v$ (or $v < u$) if the argument of u is larger than that of v in the counter-clockwise sense.

Theorem 6.2 (Hryniewicz, Momin, and Salomão [2015]). *Consider a Reeb flow on (S^3, ξ_0) that admits a pair of periodic orbits P_0, P_1 forming a Hopf link. Denote by θ_0, θ_1 their transverse rotation numbers computed with respect to Seifert surfaces. If (p, q) is a pair of relatively prime integers satisfying*

$$(\theta_0, 1) < (p, q) < (1, \theta_1) \qquad \text{or} \qquad (1, \theta_1) < (p, q) < (\theta_0, 1)$$

then there is a periodic orbit $P \subset S^3 \setminus (P_0 \cup P_1)$ such that $p = \operatorname{link}(P, P_0)$ and $q = \operatorname{link}(P, P_1)$.

The main tools in the proof are the contact homology theory introduced by Momin [2011] and the intersection theory of punctured holomorphic curves in dimension four developed by Siefring [2011].

Another source of motivation for Theorem 6.2 is a result due to Angenent [2005] which we now recall. It concerns geodesic flows on Riemannian two-spheres. Let g be a Riemannian metric on S^2, and let $\gamma : \mathbb{R} \to S^2$ be a closed geodesic of length L parametrized with unit speed. In particular $\gamma(t)$ is L-periodic. Jacobi fields along γ are characterized by the second order ODE $y''(t) = -K(\gamma(t))y(t)$ where K denotes the Gaussian curvature. Given a (non-trivial) solution $y(t)$ we can write $y'(t) + iy(t) = r(t)e^{i\theta(t)}$ in polar coordinates. The Poincaré inverse rotation number of γ is defined as

$$\rho(\gamma) = \frac{L}{2\pi} \lim_{t \to +\infty} \frac{\theta(t)}{t} \tag{17}$$

The special case of the results from Angenent [ibid.] that we would like to emphasize concerns the case when γ is simple, that is, $\gamma|_{[0,L)}$ is injective. Denote by $n(t)$ a normal vector along $\gamma(t)$. Given relatively prime integers $p, q \neq 0$ and $\epsilon > 0$ small, a (p, q)-satellite about γ is the equivalence class of the immersion $\alpha_\epsilon : \mathbb{R}/\mathbb{Z} \to S^2$

$$\alpha_\epsilon(t) = \exp_{\gamma(qtL)}\left(\epsilon \sin(2\pi pt)\, n(qtL)\right). \tag{18}$$

Two immersions are equivalent if they are homotopic through immersions, but self-tangencies and tangencies with γ are not allowed.

Theorem 6.3 (Angenent [2005]). *If a rational number p'/q' strictly between $\rho(\gamma)$ and 1 is written in lowest terms then there exists a closed geodesic which is a (p',q')-satellite about γ.*

Let us explain the connection between theorems 6.2 and 6.3. The unit tangent bundle $T^1S^2 = \{v \in TS^2 \mid g(v,v) = 1\}$ admits a contact form λ_g whose Reeb flow coincides with the geodesic flow. It is given by the restriction to T^1S^2 of the pull-back of the tautological 1-form on T^*S^2 by the associated Legendre transform. The L-periodic orbits $\dot\gamma(t)$ and $-\dot\gamma(-t)$ form a link l_γ on T^1S^2 transverse to the contact structure $\ker\lambda_g$. There exists a double cover $S^3 \to T^1S^2$ that pulls back the Reeb flow of λ_g to a Reeb flow on (S^3,ξ_0). Moreover, it pulls back the link l_γ to a Hopf link consisting of periodic orbits $P_0 \cup P_1$ just like in the statement of Theorem 6.2. Note that $\rho(\gamma) \neq 1$ forces the vectors $(\theta_0 = 2\rho(\gamma)-1, 1)$ and $(1, \theta_1 = 2\rho(\gamma)-1)$ to span a non-empty sector. Then Theorem 6.2 captures the contractible (p',q')-satellites of Theorem 6.3 up to homotopy, and a refinement for Reeb flows on the standard $\mathbb{R}P^3$ (Hryniewicz, Momin, and Salomão [2015, Theorem 1.9]) captures all the (p',q')-satellites of Theorem 6.3 up to homotopy. Of course, we do not hope to capture geodesics up to equivalence of satellites because Theorem 6.2 deals with more general flows than those dealt by Theorem 6.3. For instance, it handles non-reversible Finsler geodesic flows with a pair of closed geodesics homotopic to a pair of embedded loops through immersions without positive tangencies. In particular, it covers reversible Finsler metrics with a simple closed geodesic.

Finally, a pair of closed Reeb orbits forming a Hopf link is not known to exist in general for a Reeb flow on (S^3,ξ_0). Each of its components is unknotted, transverse to ξ_0 and has self-linking number -1; we refer to such a closed curve as a Hopf fiber. The existence of at least one closed Reeb orbit on (S^3,ξ_0) which is a Hopf fiber is proved in Hofer, Wysocki, and Zehnder [1996b]; this is a difficult result. If P is a nondegenerate closed orbit which bounds a disk-like global surface of section then P is a Hopf fiber and its rotation number is > 1. Moreover, a fixed point of the first return map, assured by Brouwer's translation theorem, determines a closed orbit P' which forms a Hopf link with P. One may ask whether every closed orbit which is a Hopf fiber and has rotation number > 1 admits another closed orbit forming together a Hopf link. In that direction we have the following result which may be seen as a version of Brouwer's translation theorem for Reeb flows on (S^3,ξ_0).

Theorem 6.4 (Hryniewicz, Momin, and Salomão [n.d.]). *Assume that a Reeb flow on (S^3,ξ_0) admits a closed Reeb orbit P which is a Hopf fiber. If the transverse rotation number $\rho(P)$ belongs to $(1,+\infty) \setminus \left\{1+\frac{1}{k} : k \in \mathbb{N}\right\}$ then there exists a closed orbit P' simply linked to P.*

The closed orbit P' in Theorem 6.4 is not even known to be unknotted.

A Closed geodesics on a Riemannian two-sphere

The purpose of this appendix is to describe the steps of a new proof of the existence of infinitely many closed geodesics on any Riemannian two-sphere. The argument is based on a combination of Angenent's theorem (Theorem 6.3) and the work of Hingston [1993], it serves as an alternative to the classical proof that combines results of Victor Bangert and John Franks. We recommend Oancea [2015] for an account of the closed geodesic problem on Riemannian manifolds.

Theorem A.1 (Bangert [1993] and Franks [1992])**.** *Every Riemannian metric on S^2 admits infinitely many closed geodesics.*

We start with a remark from Hingston [1993]. The space of embedded loops in the two-sphere carries a 3-dimensional homology class modulo short loops. One can use Grayson's curve shortening flow to run a min-max argument over this class and obtain a special simple closed geodesic γ_*. The crucial fact here is that Grayson's curve shortening flow preserves the property of being embedded. The sum of the Morse index and the nullity of γ_* is larger than or equal to 3, in particular $\rho(\gamma_*) \geq 1$.

If $\rho(\gamma_*) = 1$ (Hingston's non-rotating case) then γ_* is a very special critical point of the energy functional. The growth of Morse indices under iterations of γ_* follows a specific pattern. Index plus nullity of γ_* is equal to 3, and if γ_* is isolated then its local homology is non-trivial in degree 3. If every iterate of γ_* is isolated then the analysis of Hingston [ibid.] shows that there are infinitely many closed geodesics. If some iterate γ_* is not isolated then already there are infinitely many closed geodesics. Hence we are left with the case $\rho(\gamma_*) > 1$, which is covered by Theorem 6.3. Theorem A.1 is proved. The case $\rho(\gamma_*) > 1$ is handled independently by Theorem 6.2.

The work of Hingston [ibid.] triggered many developments, including a proof of the Conley conjecture for standard symplectic tori in Hingston [2009]. In Ginzburg [2010] used Floer homology and Hingston's methods to prove the Conley conjecture for aespherical symplectic manifolds.

References

Alberto Abbondandolo, Barney Bramham, Umberto L. Hryniewicz, and Pedro A. S. Salomão (2017a). "A systolic inequality for geodesic flows on the two-sphere". *Math. Ann.* 367.1-2, pp. 701–753. MR: 3606452 (cit. on pp. 968, 970).

Alberto Abbondandolo, Barney Bramham, Umberto L. Hryniewicz, and Pedro A. S. Salomão (Sept. 2017b). "Contact forms with large systolic ratio in dimension three". To appear in *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*. arXiv: 1709.01621 (cit. on pp. 968, 972).

– (Oct. 2017c). "Systolic ratio, index of closed orbits and convexity for tight contact forms on the three-sphere". arXiv: 1710.06193 (cit. on pp. 968, 973).

– (2018). "Sharp systolic inequalities for Reeb flows on the three-sphere". *Invent. Math.* 211.2, pp. 687–778. MR: 3748316 (cit. on pp. 968, 972).

Peter Albers, Joel W. Fish, Urs Frauenfelder, Helmut Hofer, and Otto van Koert (2012). "Global surfaces of section in the planar restricted 3-body problem". *Arch. Ration. Mech. Anal.* 204.1, pp. 273–284. MR: 2898741 (cit. on p. 975).

Peter Albers, Joel W. Fish, Urs Frauenfelder, and Otto van Koert (2013). "The Conley-Zehnder indices of the rotating Kepler problem". *Math. Proc. Cambridge Philos. Soc.* 154.2, pp. 243–260. MR: 3021812 (cit. on p. 976).

Peter Albers, Urs Frauenfelder, Otto van Koert, and Gabriel P. Paternain (2012). "Contact geometry of the restricted three-body problem". *Comm. Pure Appl. Math.* 65.2, pp. 229–263. MR: 2855545 (cit. on p. 976).

J. C. Álvarez Paiva and F. Balacheff (2014). "Contact geometry and isosystolic inequalities". *Geom. Funct. Anal.* 24.2, pp. 648–669. MR: 3192037 (cit. on p. 969).

Sigurd B. Angenent (2005). "Curve shortening and the topology of closed geodesics on surfaces". *Ann. of Math. (2)* 162.3, pp. 1187–1241. MR: 2179729 (cit. on pp. 981, 982).

Florent Balacheff (2006). "Sur la systole de la sphère au voisinage de la métrique standard". *Geom. Dedicata* 121, pp. 61–71. MR: 2276235 (cit. on p. 969).

– (2010). "A local optimal diastolic inequality on the two-sphere". *J. Topol. Anal.* 2.1, pp. 109–121. MR: 2646992 (cit. on p. 969).

Victor Bangert (1993). "On the existence of closed geodesics on two-spheres". *Internat. J. Math.* 4.1, pp. 1–10. MR: 1209957 (cit. on pp. 960, 983).

George D Birkhoff (1914). "The restricted problem of three bodies". *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 39.1, pp. 265–334 (cit. on pp. 974, 975).

George D. Birkhoff (1913). "Proof of Poincaré's geometric theorem". *Trans. Amer. Math. Soc.* 14.1, pp. 14–22. MR: 1500933 (cit. on p. 980).

– (1966). *Dynamical systems*. With an addendum by Jurgen Moser. American Mathematical Society Colloquium Publications, Vol. IX. American Mathematical Society, Providence, R.I., pp. xii+305. MR: 0209095 (cit. on p. 961).

F. Bourgeois, Y. Eliashberg, H. Hofer, K. Wysocki, and E. Zehnder (2003). "Compactness results in symplectic field theory". *Geom. Topol.* 7, pp. 799–888. MR: 2026549.

B. Bramham and H. Hofer (2012). "First steps towards a symplectic dynamics". In: *Surveys in differential geometry. Vol. XVII*. Vol. 17. Surv. Differ. Geom. Int. Press, Boston, MA, pp. 127–177. MR: 3076060 (cit. on p. 960).

Barney Bramham (2015a). "Periodic approximations of irrational pseudo-rotations using pseudoholomorphic curves". *Ann. of Math. (Second Series)* 181.3, pp. 1033–1086. MR: 3296820 (cit. on p. 979).

– (2015b). "Pseudo-rotations with sufficiently Liouvillean rotation number are C^0-rigid". *Invent. Math.* 199.2, pp. 561–580. MR: 3302121 (cit. on p. 979).

Christopher B. Croke (1988). "Area and the length of the shortest closed geodesic". *J. Differential Geom.* 27.1, pp. 1–21. MR: 918453 (cit. on p. 968).

Joel W. Fish and Richard Siefring (Nov. 2013). "Connected sums and finite energy foliations I: Contact connected sums". arXiv: 1311.4221 (cit. on p. 979).

John Franks (1992). "Geodesics on S^2 and periodic points of annulus homeomorphisms". *Invent. Math.* 108.2, pp. 403–418. MR: 1161099 (cit. on pp. 960, 964, 983).

David Fried (1982). "The geometry of cross sections to flows". *Topology* 21.4, pp. 353–371. MR: 670741 (cit. on p. 961).

Hansjörg Geiges (2008). *An introduction to contact topology*. Vol. 109. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, pp. xvi+440. MR: 2397738 (cit. on p. 962).

Étienne Ghys (2009). "Right-handed vector fields & the Lorenz attractor". *Jpn. J. Math.* 4.1, pp. 47–61. MR: 2491282 (cit. on p. 961).

Viktor L. Ginzburg (2010). "The Conley conjecture". *Ann. of Math. (2)* 172.2, pp. 1127–1180. MR: 2680488 (cit. on p. 983).

Emmanuel Giroux (2002). "Géométrie de contact: de la dimension trois vers les dimensions supérieures". In: *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*. Higher Ed. Press, Beijing, pp. 405–414. MR: 1957051 (cit. on p. 965).

Mikhael Gromov (1983). "Filling Riemannian manifolds". *J. Differential Geom.* 18.1, pp. 1–147. MR: 697984 (cit. on p. 968).

– (1996). "Systoles and intersystolic inequalities". In: *Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992)*. Vol. 1. Sémin. Congr. Soc. Math. France, Paris, pp. 291–362. MR: 1427763.

Adam Harris and Gabriel P. Paternain (2008). "Dynamically convex Finsler metrics and J-holomorphic embedding of asymptotic cylinders". *Ann. Global Anal. Geom.* 34.2, pp. 115–134. MR: 2425525 (cit. on pp. 966, 967).

Nancy Hingston (1993). "On the growth of the number of closed geodesics on the two-sphere". *Internat. Math. Res. Notices* 9, pp. 253–262. MR: 1240637 (cit. on pp. 960, 983).

– (2009). "Subharmonic solutions of Hamiltonian equations on tori". *Ann. of Math. (2)* 170.2, pp. 529–560. MR: 2552101 (cit. on p. 983).

H. Hofer (1993). "Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three". *Invent. Math.* 114.3, pp. 515–563. MR: 1244912 (cit. on p. 960).

H. Hofer, K. Wysocki, and E. Zehnder (1995a). "A characterisation of the tight three-sphere". *Duke Math. J.* 81.1. A celebration of John F. Nash, Jr., 159–226 (1996). MR: 1381975 (cit. on pp. 966, 967, 978).

– (1995b). "Properties of pseudo-holomorphic curves in symplectisations. II. Embedding controls and algebraic invariants". *Geom. Funct. Anal.* 5.2, pp. 270–328. MR: 1334869.

– (1996a). "Properties of pseudoholomorphic curves in symplectisations. I. Asymptotics". *Ann. Inst. H. Poincaré Anal. Non Linéaire* 13.3, pp. 337–379. MR: 1395676.

– (1996b). "Unknotted periodic orbits for Reeb flows on the three-sphere". *Topol. Methods Nonlinear Anal.* 7.2, pp. 219–244. MR: 1481697 (cit. on pp. 963, 982).

– (1998). "The dynamics on three-dimensional strictly convex energy surfaces". *Ann. of Math. (2)* 148.1, pp. 197–289. MR: 1652928 (cit. on pp. 961, 963, 964, 973, 975, 978, 980).

– (1999a). "A characterization of the tight 3-sphere. II". *Comm. Pure Appl. Math.* 52.9, pp. 1139–1177. MR: 1692144 (cit. on pp. 966, 967, 973, 978).

– (1999b). "Properties of pseudoholomorphic curves in symplectizations. III. Fredholm theory". In: *Topics in nonlinear analysis*. Vol. 35. Progr. Nonlinear Differential Equations Appl. Birkhäuser, Basel, pp. 381–475. MR: 1725579.

– (2003). "Finite energy foliations of tight three-spheres and Hamiltonian dynamics". *Ann. of Math. (2)* 157.1, pp. 125–255. MR: 1954266 (cit. on pp. 960, 962, 977–979).

Helmut Hofer and Eduard Zehnder (1994). *Symplectic invariants and Hamiltonian dynamics*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel, pp. xiv+341. MR: 1306732 (cit. on p. 970).

Umberto L. Hryniewicz (2012). "Fast finite-energy planes in symplectizations and applications". *Trans. Amer. Math. Soc.* 364.4, pp. 1859–1931. MR: 2869194 (cit. on pp. 962, 963, 978).

– (2014). "Systems of global surfaces of section for dynamically convex Reeb flows on the 3-sphere". *J. Symplectic Geom.* 12.4, pp. 791–862. MR: 3333029 (cit. on pp. 962, 963, 978).

Umberto L. Hryniewicz, Joan E. Licata, and Pedro A. S. Salomão (2015). "A dynamical characterization of universally tight lens spaces". *Proc. Lond. Math. Soc. (3)* 110.1, pp. 213–269. MR: 3299604 (cit. on pp. 964, 967, 978).

Umberto L. Hryniewicz, A. Momin, and Pedro A. S. Salomão (n.d.). "On simply linked Reeb orbits in the tight three-sphere". In preparation (cit. on p. 982).

Umberto L. Hryniewicz, Al Momin, and Pedro A. S. Salomão (2015). "A Poincaré-Birkhoff theorem for tight Reeb flows on S^3". *Invent. Math.* 199.2, pp. 333–422. MR: 3302117 (cit. on pp. 960, 981, 982).

Umberto L. Hryniewicz and Pedro A. S. Salomão (2011). "On the existence of disk-like global sections for Reeb flows on the tight 3-sphere". *Duke Math. J.* 160.3, pp. 415–465. MR: 2852366 (cit. on pp. 964, 978).

– (2013). "Global properties of tight Reeb flows with applications to Finsler geodesic flows on S^2". *Math. Proc. Cambridge Philos. Soc.* 154.1, pp. 1–27. MR: 3002580 (cit. on p. 967).

– (2016). "Elliptic bindings for dynamically convex Reeb flows on the real projective three-space". *Calc. Var. Partial Differential Equations* 55.2, Art. 43, 57. MR: 3485982 (cit. on p. 966).

Umberto L. Hryniewicz, Pedro A. S. Salomão, and K. Wysocki (n.d.). "Genus zero global surfaces of section for Reeb flows in dimension three". In preparation (cit. on pp. 965, 966).

Al Momin (2011). "Contact homology of orbit complements and implied existence". *J. Modern Dyn.* 5.4, pp. 409–472 (cit. on p. 981).

Alexandru Oancea (2015). "Morse theory, closed geodesics, and the homology of free loop spaces". In: *Free loop spaces in geometry and topology*. Vol. 24. IRMA Lect. Math. Theor. Phys. With an appendix by Umberto Hryniewicz. Eur. Math. Soc., Zürich, pp. 67–109. MR: 3444362 (cit. on p. 983).

Naiara V. de Paulo and Pedro A. S. Salomão (n.d.). "On the multiplicity of periodic orbits and homoclinics near critical energy levels of Hamiltonian systems in $\mathbb{R}^4$". To appear in *Transactions of the Amer. Math. Soc.* (cit. on p. 979).

– (2018). "Systems of transversal sections near critical energy levels of Hamiltonians systems in $\mathbb{R}^4$". *Memoirs of the Amer. Math. Soc.* 252.1202 (cit. on p. 979).

Henri Poincaré (1912). "Sur un théorème de géométrie". *Rend. Circ. Mat. Palermo* 33, pp. 375–407 (cit. on pp. 959, 980).

Hans-Bert Rademacher (2007). "Existence of closed geodesics on positively curved Finsler manifolds". *Ergodic Theory Dynam. Systems* 27.3, pp. 957–969. MR: 2322187 (cit. on p. 966).

Pedro A. S. Salomão (2003). "Convex energy levels of Hamiltonian systems". *Qual. Theory Dyn. Syst.* 4.2, 439–457 (2004). MR: 2129729 (cit. on p. 979).

Sol Schwartzman (1957). "Asymptotic cycles". *Ann. of Math. (2)* 66, pp. 270–284. MR: 0088720 (cit. on p. 961).

Richard Siefring (2011). "Intersection theory of punctured pseudoholomorphic curves". *Geom. Topol.* 15.4, pp. 2351–2457. MR: 2862160 (cit. on pp. 965, 981).

Dennis Sullivan (1976). "Cycles for the dynamical study of foliated manifolds and complex manifolds". *Invent. Math.* 36, pp. 225–255. MR: 0433464 (cit. on p. 961).

Claude Viterbo (2000). "Metric and isoperimetric problems in symplectic geometry". *J. Amer. Math. Soc.* 13.2, pp. 411–431. MR: 1750956 (cit. on pp. 968, 970).

Received 2018-02-07.

PEDRO A. S. SALOMÃO
psalomao@ime.usp.br

UMBERTO L. HRYNIEWICZ
umbertolh@gmail.com

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (987–1010)

STABILITY CONDITIONS IN SYMPLECTIC TOPOLOGY

Ivan Smith

Abstract

We discuss potential (largely speculative) applications of Bridgeland's theory of stability conditions to symplectic mapping class groups.

1 Introduction

A symplectic manifold (X, ω) which is closed or convex at infinity has a Fukaya category $\mathcal{F}(X, \omega)$, which packages the algebraic information held by moduli spaces of holomorphic discs in X with Lagrangian boundary conditions Seidel [2008] and Fukaya, Oh, Ohta, and Ono [2009]. The associated category of perfect modules $D^\pi \mathcal{F}(X, \omega) = \mathcal{F}(X, \omega)^{perf}$ is a triangulated category, linear over the Novikov field Λ. The category is $\mathbb{Z}$-graded whenever $2c_1(X) = 0$.

Any $\mathbb{Z}$-graded triangulated category $\mathcal{C}$ has an associated complex manifold $\mathrm{Stab}(\mathcal{C})$ of stability conditions Bridgeland [2007], which carries an action of the group $\mathrm{Auteq}(\mathcal{C})$ of triangulated autoequivalences of $\mathcal{C}$. Computation of the space of stability conditions remains challenging, but a number of instructive examples are now available, and it is reasonable to wonder what the theory of stability conditions might say about symplectic topology. One direction in which one can hope for non-trivial applications concerns global structural features of the symplectic mapping class group. The existing theory will need substantial development for this to bear fruit, so a survey risks being quixotic, but it provides a vantage-point from which to interpret some recent activity Bridgeland and Smith [2015], Smith [2015], and Sheridan and Smith [2017b].

Partially funded by a Fellowship from EPSRC. Thanks to friends and colleagues at Cambridge for a decade and a half of moral support, intellectual stimulation and company lifting spirits.
MSC2010: primary 53D35; secondary 53D37, 18E30, 57R50.

Acknowledgments. These notes draw on discussions or joint work with all of Denis Auroux, Arend Bayer, Tom Bridgeland, Jonny Evans, Daniel Huybrechts, Oscar Randal-Williams, Tony Scholl, Paul Seidel, Nick Sheridan and Michael Wemyss. More generally, I am extremely indebted to all my collaborators, who have taught me more than I (and probably they) would have thought possible.

2 Mapping class groups

2.1 Surfaces. Consider a closed oriented surface Σ_g of genus g. The mapping class group $\Gamma_g = \pi_0 \mathrm{Diff}(\Sigma_g)$ has been the focus of an enormous amount of attention, from many viewpoints. Here are a few of its salient features, see Farb and Margalit [2012], Calegari [2009], and Morita [2007] for detailed references.

1. Γ_g is finitely presented. The action on homology defines a natural surjection $\Gamma_g \to Sp_{2g}(\mathbb{Z})$ with kernel the Torelli group I_g; the latter is torsion-free, infinitely generated when $g = 2$, finitely generated (but not known to be finitely presented) when $g > 2$.

2. Γ_g has finite rational cohomological dimension. It has finitely many conjugacy classes of finite subgroups; is generated by finitely many torsion elements; satisfies the Tits alternative; and is residually finite, in particular contains no non-trivial divisible element.

3. There is a dynamical classification of mapping classes: given simple closed curves α, β on Σ, the geometric intersection numbers $\iota(\alpha, \phi^n(\beta))$ are either periodic, grow linearly or grow exponentially. A random walk on the mapping class group will almost surely end on the last case. Γ_g has non-trivial quasi-morphisms (and scl), hence does not obey a law.

Here scl denotes stable commutator length. Recall that a group G *obeys a law* if there is a word w in a free group $\mathbb{F}$ for which every homomorphism $\mathbb{F} \to G$ sends $w \mapsto \mathrm{id}$. (Abelian, nilpotent and solvable groups obey laws; groups with non-vanishing scl do not.) It is worth pointing out that plenty of elementary things remain unknown: must a subgroup of I_g generated by two elements be commutative or free?

Broadly speaking, the above results might be divided into three categories: results (finite presentability, non-existence of non-trivial divisible elements, etc) which underscore the similarities between the mapping class group and arithmetic groups; results of a general group-theoretic or finiteness nature, in the vein of geometric group theory; and results (scl, behaviour of random walks) of a more dynamical flavour, in some cases connecting via geometric intersection numbers rather directly to Floer theory. These three directions suggest broad classes of question one might ask about the (symplectic) mapping class groups of higher-dimensional manifolds.

Much of what we know about Γ_g comes from the fact that, although not a hyperbolic group when $g > 1$ (it contains free abelian subgroups of rank $2g - 1 > 1$), it acts on geometrically meaningful spaces which are non-positively-curved in a suitable sense: Teichmüller space, on the one hand, and the complex of curves on another (the latter is δ-hyperbolic Masur and Minsky [1999]). Indeed, Γ_g can be characterised via these actions, as the isometry group of the former for the Teichmüller metric (when $g > 2$; for $g = 2$ one should quotient by the hyperelliptic involution) Royden [1970], or as the simplicial isometry group of the latter Ivanov [1997].

2.2 Higher-dimensional smooth manifolds. There is no simply-connected manifold of dimension > 3 for which the homotopy type of the diffeomorphism group $\mathrm{Diff}(M)$ is known completely. Nonetheless, there are many broad structural results concerning diffeomorphism groups and mapping class groups in high dimensions. Deep results in surgery theory Sullivan [1977] imply that if M is a simply-connected manifold of dimension at least five, then $\pi_0\mathrm{Diff}(M)$ is commensurable with an arithmetic group, hence is finitely presented and contains no non-trivial divisible elements. The arithmetic group arises from automorphisms of the Sullivan minimal model; the forgetful map from $\pi_0\mathrm{Diff}(M)$ to the group of tangential self-homotopy equivalences of M has finite kernel up to conjugacy Browder [1967]. Arithmetic groups cannot contain non-trivial divisible elements. Indeed, their eigenvalues, in a fixed matrix representation, are algebraic integers of bounded degree. Therefore a divisible subgroup is composed of unipotent elements, and has Zariski closure a unipotent subgroup; but commutative unipotent groups only have free abelian arithmetic subgroups Kambayashi, Miyanishi, and Takeuchi [1974]. (Note that general finitely presented groups *can* contain non-trivial divisible elements: indeed, every finitely presented group is a subgroup of some finitely presented group G whose commutator subgroup $[G, G]$ has only divisible elements and with $G/[G, G] \cong \mathbb{Z}$, cf. Baumslag and Miller [1988].) The same conclusions, in particular finite presentability, also apply to the Torelli group.

Away from the simply connected case, much less is known. For $n \geqslant 6$, the mapping class group $\pi_0\mathrm{Diff}(T^n)$ has a split subgroup $(\mathbb{Z}/2)^\infty$ arising from the Whitehead group Wh_2 of the fundamental group Hatcher [1978] and Hsiang and Sharpe [1976], so finite generation can fail. (This is in fact a phenomenon about homeomorphisms, rather distinct from questions of exotic smooth structures.) The group of homotopy self-equivalences $hAut(M)$ when $\pi_1(M)$ is general also seems mysterious: can $hAut(M)$, or $\pi_0\mathrm{Diff}(M)$, ever contain a divisible group, for instance the rationals, as a subgroup? Given the lack of complete computations, nothing seems to be known about the realisability problem, i.e.

which groups occur as smooth mapping class groups: is there an M for which $\pi_0\mathrm{Diff}(M)$ is a rank 2 free group?

2.3 Symplectic manifolds. Fix a symplectic $2n$-manifold (X,ω), closed or convex at infinity. Let $\mathrm{Symp}(X,\omega)$ denote the group of diffeomorphisms of X preserving ω, equipped with the C^∞-topology.

1. If $H^1(X;\mathbb{R}) = 0$ then the connected component of the identity $\mathrm{Symp}_0(X) = \mathrm{Symp}(X,\omega)\cap\mathrm{Diff}_0(X)$ is exactly the subgroup $\mathrm{Ham}(X)$ of Hamiltonian symplectomorphisms, and the symplectic mapping class group is the quotient $\pi_0\mathrm{Symp}(X) = \mathrm{Symp}(X,\omega)/\mathrm{Ham}(X)$.
2. If $H^1(X;\mathbb{R}) \neq 0$ then there is a flux homomorphism (defined on the universal cover)
$$\Phi : \widetilde{\mathrm{Symp}}(X,\omega) \to H^1(X;\mathbb{R})$$
whose kernel is the universal cover of the Hamiltonian group. The image Γ of Φ viewed as a homomorphism on $\pi_1\mathrm{Symp}(X,\omega)$ is a discrete group Ono [2006], and $\mathrm{Symp}_0(X)/\mathrm{Ham}(X) \cong H^1(X;\mathbb{R})/\Gamma$. $\mathrm{Symp}_0(X,\omega)$ carries a foliation by copies of $\mathrm{Ham}(X)$, whose leaf space is locally isomorphic to $H^1(X;\mathbb{R})$. Since Floer theory is Hamiltonian invariant, it is sometimes useful to equip $\mathrm{Symp}(X,\omega)$ with the weaker "Hamiltonian topology", in which (by definition) only isotopies along the leaves are continuous.

There is a forgetful map
$$q : \pi_0\mathrm{Symp}(X,\omega) \to \pi_0\mathrm{Diff}(X)$$
which is typically neither injective nor surjective. For non-injectivity, one has Seidel's results on squared Dehn twists and their inheritors Seidel [1999] and Tonkonog [2015]: if $L \subset X$ is a Lagrangian sphere, there is an associated twist τ_L, which has finite order smoothly if $n = \dim_{\mathbb{R}}(L)$ is even, but typically has infinite order symplectically, for instance for smooth hypersurfaces of degree > 2 in projective space. If $n = 2$, the smooth finiteness can be seen rather explicitly. The cotangent bundle $T^*S^2 = \{x^2+y^2+z^2 = 1\} \subset \mathbb{C}^3$ is an affine quadric; the Dehn twist τ_L in the zero-section L is the monodromy of the Lefschetz degeneration $\{x^2+y^2+z^2 = t\}$ of that quadric around the unit circle in the t-plane; that family base-changes to give a family with a 3-fold node $\{x^2+y^2+z^2 = t^2\}$; and the node admits a small resolution, so the base-change family has trivial smooth monodromy and $\tau_L^2 = \mathrm{id}$ in (compactly supported) C^∞.

For non-surjectivity, a diffeomorphism cannot be isotopic to a symplectic diffeomorphism unless it preserves $[\omega]$ and $c_i(X) \in H^{2i}(X;\mathbb{Z})$, and furthermore permutes the classes with non-trivial Gromov-Witten invariants Ruan [1993]. There are also diffeomorphisms acting trivially on cohomology which do not preserve the homotopy class of any almost complex structure Randal-Williams [2015], so q is not onto the Torelli subgroup for many projective hypersurfaces.

There should be further, less homotopy-theoretic constraints on Torelli symplectomorphisms. A diffeomorphism $\phi : X \to X$ induces an automorphism of the A_∞-algebra $H^*(X;\mathbb{C})$ (where the A_∞-structure is the classical one on cohomology) which may have non-trivial higher-order terms even when the linear action is trivial. If X is Fano, its quantum cohomology

$$QH^*(X;\mathbb{C}) \cong \oplus_\lambda QH^*(X;\lambda) \qquad \lambda \in \mathrm{Spec}(*c_1(X)) \tag{2-1}$$

splits into generalised eigenspaces for quantum product by $c_1(X)$, and the higher-order terms of any symplectomorphism should preserve this decomposition, i.e. be compatible with the quantum-corrected A_∞-structure, a non-linear constraint.

2.4 Monodromy. If X admits a Kähler metric g with positive-dimensional isometry group $K \subset \mathrm{Isom}(X,g)$, then the map $K \to \mathrm{Symp}(X,\omega)$ is often rather interesting, at least for higher homotopy groups. Gromov [1985] proved that $\mathbb{P}U(3) \simeq \mathrm{Symp}(\mathbb{P}^2)$ and inferred that the group of compactly supported symplectomorphisms $\mathrm{Symp}_{ct}(\mathbb{C}^2)$ is contractible; for $n > 2$ we don't know anything about $\pi_0\mathrm{Symp}_{ct}(\mathbb{C}^n,\omega_{\mathrm{st}})$. It follows that fully computing $\pi_0\mathrm{Symp}(X,\omega)$ will not be feasible in essentially any higher-dimensional case, and replacing it with an algebraic avatar makes sense.

The most obvious source of information regarding $\pi_0\mathrm{Symp}(X,\omega)$ comes from algebraic geometry: if X is a smooth projective variety and varies in a moduli space $\mathcal{M}$ (of complex structures, or complete intersections, with fixed polarisation), parallel transport yields a map

$$\rho : \pi_1(\mathcal{M}) \to \pi_0\mathrm{Symp}(X,\omega); \tag{2-2}$$

most known constructions of non-trivial symplectic mapping classes arise this way. Even in simple cases, classical topology struggles to say much about (2-2). Let p be a degree m polynomial with distinct roots. Let X be the Milnor fibre $\{x^2+y^2+p(z)=0\} \subset \mathbb{C}^3$ of the A_{m-1}-singularity $\mathbb{C}^2/(\mathbb{Z}/m)$. Parallel transport for the family over configuration space varying p, and simultaneous resolution of 3-fold nodes, now yields a diagram

$$\begin{array}{ccc} Br_m & \xrightarrow{\ \rho\ } & \pi_0\mathrm{Symp}_{ct}(X) \\ \downarrow & & \downarrow \\ \mathrm{Sym}_m & \xrightarrow{\ \mu\ } & \pi_0\mathrm{Diff}_{ct}(X) \end{array} \tag{2-3}$$

but ρ is actually injective Khovanov and Seidel [2002]. For Milnor fibres of most singularities, the kernel

$$\pi_0\mathrm{Symp}_{ct}(X,\omega) \to \pi_0\mathrm{Diff}_{ct}(X)$$

is large Keating [2014]. In another direction, let $\mathcal{M}_{d,n} = Z_{d,n}/\mathbb{P}\,\mathrm{GL}_{n+2}(\mathbb{C})$ denote the moduli space of degree d hypersurfaces in $\mathbb{P}^{n+1}$, with $Z_{d,n} \subset H^0(\mathbb{P}^{n+1}, \mathcal{O}(d))$ the discriminant complement. Tommasi [2014] proved that $\tilde{H}^*(\mathcal{M}_{d,n};\mathbb{Q}) = 0$ for $d \geqslant 3$ and $* < (d+1)/2$, so the natural map $\mathcal{M}_{d,n} \to B\mathrm{Diff}(X_{d,n})$ cannot be probed using the rather rich cohomology of the image Galatius and O. Randal-Williams [2017]. It seems unclear when (2-2) induces an interesting map on rational cohomology.

Whilst (the possible failure of) *injectivity* of ρ can be probed using Floer-theoretic methods, these seem much less well-adapted to understanding the possible *surjectivity* of ρ; for questions of finite generation, or residual finiteness, one needs maps out of $\pi_0\mathrm{Symp}(X,\omega)$, i.e. one needs to have it act on something. Whilst one can formally write down analogues of the complex of curves, the lack of classification results for Lagrangian submanifolds makes it hard to extract useful information. Spaces of stability conditions on the Fukaya category emerge as another contender, simply because they have been effective in some situations on the other side of the mirror.

Obviously, since $\pi_0\mathrm{Symp}(X,\omega)$ acts on spaces associated to the Fukaya category through the action of the quotient $\mathrm{Auteq}(\mathcal{F}(X,\omega))/\langle[2]\rangle$ (dividing by twice the shift functor), one can only hope to extract information about the latter; this is somewhat similar to first steps in surgery theory, in which one obtains information about $\pi_0\mathrm{Diff}(M)$ from the quotient group $hAut(M)$ of homotopy self-equivalences on the one-hand, as probed by surgery theory and L-theory, and separately (and with separate techniques) from pseudo-isotopy theory and algebraic K-theory. In classical surgery theory, $B\mathrm{Diff}(M)$ is more accessible to study (e.g. cohomologically) than $\pi_0\mathrm{Diff}(M)$, and working at the space level is crucial for many applications. One can upgrade autoequivalences of an A_∞-category to a simplicial space, but there are no compelling computations, and we will not pursue that direction. We should at least mention the Seidel representation Seidel [1997] $\pi_1\mathrm{Ham}(X,\omega) \to QH^*(X)^\times$ (to the group of invertibles) as indication that the higher homotopy groups may carry interesting information away from the Calabi-Yau setting; when $c_1(X) = 0$ the situation is less clear-cut.

We will say that a symplectic manifold (X,ω) is *homologically mirror* to an algebraic variety X° over Λ if there is a Λ-linear equivalence $D^\pi\mathcal{F}(X,\omega) \simeq D^b(X^\circ)$. The following result, whilst narrow in scope, gives a first indication that knowing that something is a homological mirror might sometimes be leveraged to extract new information.

Proposition 2.1. *Let (X,ω) be a $K3$ surface which is homologically mirror to an algebraic $K3$ surface X°. Then a symplectomorphism f of X which preserves the Lagrangian* *isotopy class of each Lagrangian sphere in X acts on $D^\pi\mathcal{F}(X)$ with finite order. If f acts on $D^\pi\mathcal{F}(X)$ non-trivially then it acts on $H^*(X;\mathbb{C})$ non-trivially.*

Proof. An autoequivalence of $D^b(X^\circ)$ acting trivially on the set of spherical objects in $D^b(X^\circ)$ arises from a transcendental automorphism of X°, cf. Huybrechts [2012, Appendix A]. The group of such is a finite cyclic group Huybrechts [2016a, Ch.3 Corollary 3.4] acting faithfully on cohomology. Finally, all spherical objects in $D^\pi\mathcal{F}(X)$ are quasi-isomorphic to Lagrangian sphere vanishing cycles Sheridan and Smith [2017b] (the proof given in *op. cit.*, relying on constraining lattice self-embeddings via discriminant considerations, can be generalised away from the case when X° has Picard rank one). □

One could view this as a weak version of the fact that Γ_g acts faithfully on the complex of curves. Note that for a double plane $X \to \mathbb{P}^2$ branched over a sextic one expects that every Lagrangian sphere arises from a degeneration of the branch locus, and the covering involution reverses orientation but preserves its (unoriented) Lagrangian isotopy class, suggesting the conclusion may be optimal.

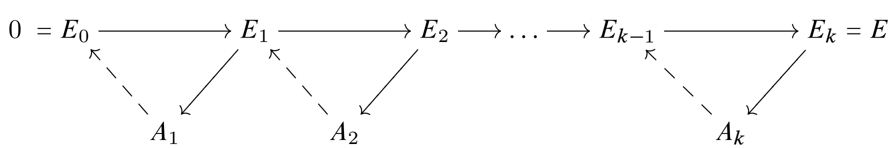
3 Stability conditions

3.1 Definitions. Let $\mathcal{C}$ be a proper, i.e. cohomologically finite, triangulated category, linear over a field k. We will assume that the numerical Grothendieck group $K(\mathcal{C})$, i.e. the quotient of the Grothendieck group $K^0(\mathcal{C})$ by the kernel of the Euler form, is free and of finite rank d. The space of (locally finite numerical) stability conditions will then be a d-dimensional complex manifold.

A *stability condition* $\sigma = (Z, \mathcal{P})$ on $\mathcal{C}$ consists of a group homomorphism $Z\colon K(\mathcal{C}) \to \mathbb{C}$ called the *central charge*, and full additive subcategories $\mathcal{P}(\phi) \subset \mathcal{C}$ of *σ-semistable objects of phase* ϕ for each $\phi \in \mathbb{R}$, which together satisfy a collection of axioms, the most important being:

(a) if $E \in \mathcal{P}(\phi)$ then $Z(E) \in \mathbb{R}_{>0} \cdot e^{i\pi\phi} \subset \mathbb{C}$;

(b) for each nonzero object $E \in \mathcal{C}$ there is a finite sequence of real numbers $\phi_1 > \phi_2 > \cdots > \phi_k$ and a collection of triangles



with $A_j \in \mathcal{P}(\phi_j)$ for all j.

We will always furthermore impose the "support property": for a norm $\|\cdot\|$ on $K(\mathcal{C})\otimes\mathbb{R}$ there is a constant $C > 0$ such that $\|\gamma\| < C\cdot|Z(\gamma)|$ for all $\gamma\in K(\mathcal{C})$ represented by σ-semistable objects in $\mathcal{C}$. (This in particular means all our stability conditions are "full", cf. Bayer and Macrì [2011, Proposition B.4].) The semistable objects A_j appearing in the filtration of axiom (b) are unique up to isomorphism, and are called the *semistable factors* of E. We set

$$\phi^+(E)=\phi_1,\quad \phi^-(E)=\phi_k,\quad m(E)=\sum_i |Z(A_i)|\in\mathbb{R}_{>0}.$$

A simple object E of $\mathcal{P}(\phi)$ is said to be stable of phase ϕ and mass $|Z(E)|$. Let $\mathrm{Stab}(\mathcal{C})$ denote the set of all stability conditions on $\mathcal{C}$. It carries a natural topology, induced by the metric

(3-1)
$$d(\sigma_1,\sigma_2)=\sup_{0\neq E\in\mathcal{C}}\left\{|\phi^-_{\sigma_2}(E)-\phi^-_{\sigma_1}(E)|,\ |\phi^+_{\sigma_2}(E)-\phi^+_{\sigma_1}(E)|,\ \left|\log\frac{m_{\sigma_2}(E)}{m_{\sigma_1}(E)}\right|\right\}\in[0,\infty].$$

Theorem 3.1 (Bridgeland [2007]). *The space* $\mathrm{Stab}(\mathcal{C})$ *has the structure of a complex manifold, such that the forgetful map* $\pi:\ \mathrm{Stab}(\mathcal{C})\longrightarrow\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}),\mathbb{C})$ *taking a stability condition to its central charge is a local isomorphism.*

The group of triangulated autoequivalences $\mathrm{Auteq}(\mathcal{C})$ acts on $\mathrm{Stab}(\mathcal{C})$ by holomorphic automorphisms which preserve the metric. There is a commuting (continuous, non-holomorphic) action of the universal cover of the group $\mathrm{GL}^+(2,\mathbb{R})$, not changing the subcategories $\mathcal{P}$, but acting by post-composition on the central charge viewed as a map to $\mathbb{C}=\mathbb{R}^2$ (and correspondingly adjusting the ϕ-labelling of $\mathcal{P}$). A subgroup $\mathbb{C}\subset\widetilde{\mathrm{GL}}^+(2,\mathbb{R})$ acts freely, by

$$\mathbb{C}\ni t:(Z,P)\mapsto(Z',\mathcal{P}'),\quad Z'(E)=e^{-i\pi t}\cdot Z(E),\quad \mathcal{P}'(\phi)=\mathcal{P}(\phi+\mathrm{Re}(t)).$$

For any integer n, the action of the shift $[n]$ coincides with the action of $n\in\mathbb{C}$. Of particular relevance is the quotient $\mathrm{Stab}(\mathcal{C})/\langle[2]\rangle$, on which the $\widetilde{\mathrm{GL}}^+(2,\mathbb{R})$-action descends to a $\mathrm{GL}^+(2,\mathbb{R})$-action; in the symplectic setting, this will amount to focussing attention on symplectomorphisms rather than graded symplectomorphisms. We will write $\mathrm{Stab}(\mathcal{F}(X))$ for $\mathrm{Stab}(D^\pi\mathcal{F}(X,\omega))$.

Remark 3.2. *If a power of the shift functor of* $\mathcal{C}$ *is isomorphic to the identity,* $\mathrm{Stab}(\mathcal{C})=\varnothing$, *so* $\mathrm{Stab}(\mathcal{F}(X))$ *is only non-trivial when* $2c_1(X)=0$. *Phantom subcategories of derived categories of coherent sheaves Gorchinskiy and Orlov [2013] have trivial K-theory so also cannot admit any stability condition.*

The definition was motivated by ideas in string theory which in turn connect closely to symplectic topology: if (X,ω) is a symplectic manifold with $c_1(X)=0$, there is conjecturally an injection from the Teichmüller space $\mathfrak{M}_X(J,\Omega)$ of marked pairs (J,Ω) comprising a compatible integrable complex structure J and J-holomorphic volume form $\Omega \in \Omega^{n,0}(X;J)$ into $\mathrm{Stab}(\mathcal{F}(X))$. In this scenario, given (J,Ω), then $Z(L) = \int_L \Omega$, the categories $\mathcal{P}(\phi)$ contain the special Lagrangians of phase ϕ, and the Harder-Narasimhan filtration should be the output of some version of mean curvature flow with surgeries at finite-time singularities, cf. Joyce [2015].

An important point is that even if one can build a map $\mathfrak{M}_X(J,\Omega) \to \mathrm{Stab}(\mathcal{F}(X))$, simply for dimension reasons it often can't be onto an open subset; one expects it to have image a complicated transcendental submanifold of high codimension. There is rarely a predicted geometric interpretation of the "general" stability condition. (Completeness of (3-1) was proven in Woolf [2012]; contrast with the Weil-Petersson metric on moduli spaces of Calabi-Yau's, which frequently has the boundary at finite distance Wang [1997] and is incomplete.) More positively, since the vanishing cycle of any nodal degeneration can be realised by a special Lagrangian Hein and Sun [2017], one expects any Lagrangian sphere which is such a vanishing cycle to be stable for some stability condition, and to define an "end" to the space of stability conditions, where the mass of this stable object tends to zero. Thus one might hope that the global topology of $\mathrm{Stab}(\mathcal{F}(X))/\mathrm{Auteq}(\mathcal{F}(X))$ carries information about Lagrangian spheres.

3.2 Properties. $\mathrm{Stab}(\mathcal{C})$ is a complex manifold. But it inherits additional structure: it is modelled on a fixed vector space such that the transition maps between charts are locally the identity, and it has a global étale map to $\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}),\mathbb{C}) \cong \mathbb{C}^d$.

Fix a connected component $\mathrm{Stab}_\dagger(\mathcal{C})$ of $\mathrm{Stab}(\mathcal{C})$. Let $\mathrm{Auteq}_\dagger(\mathcal{C})$ denote the quotient of the subgroup of $\mathrm{Auteq}(\mathcal{C})$ which preserves $\mathrm{Stab}_\dagger(\mathcal{C})$ by the subgroup of "negligible" autoequivalences, i.e. those which act trivially on $\mathrm{Stab}_\dagger(\mathcal{C})$. Negligible autoequivalences can exist, for instance coming from automorphisms of projective surfaces which act trivially on the algebraic cohomology (these need not be trivial on the transcendental cohomology).

Lemma 3.3. $\mathrm{Stab}_\dagger(\mathcal{C})$ *has a well-defined integral affine structure and a canonical measure. The quotient* $\mathcal{Q} = \mathrm{Stab}_\dagger(\mathcal{C})/\mathrm{Auteq}_\dagger(\mathcal{C})$ *is a complex orbifold, with* $2c_i(\mathcal{Q}) = 0$ *for* i *odd.*

Proof. The affine structure comes from the integral lattice in $K(\mathcal{C})$. The measure is obtained from Lebesgue measure in $\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}),\mathbb{C})$, normalised so that the quotient torus $\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}),\mathbb{C}/(\mathbb{Z}\oplus i\mathbb{Z}))$ has volume one. We claim that, having killed the generic stabiliser, $\mathrm{Auteq}_\dagger(\mathcal{C})$ acts with finite stabilisers. The argument is due to Bridgeland, and applies to the space of stability conditions satisfying the support condition whenever the central charge is required to factor through a finite rank lattice N (see also Huybrechts [2016b]). Given any stability condition σ and $R \in \mathbb{R}_{>0}$, there are only finitely many classes $n \in N$ which are represented by a σ-semistable object E with $|Z_\sigma(E)| < R$. For

$R \gg 0$ sufficiently large, these classes will span the lattice (fix a set of objects of $\mathcal{C}$ whose classes give a basis of N, and consider the classes represented by the semistable factors in the HN-filtrations of those objects). Any autoequivalence ϕ fixing σ must permute this finite set, so some fixed power σ^k fixes all these elements pointwise. Then σ^k acts trivially on $\mathrm{Hom}_{\mathbb{Z}}(N, \mathbb{C})$, hence by Theorem 3.1 on a neighbourhood of $\sigma \in \mathrm{Stab}_{\dagger}(\mathcal{C})$, and hence acts trivially globally (since it's a holomorphic automorphism). This shows that for any $\sigma \in \mathrm{Stab}_{\dagger}(\mathcal{C})$, the stabiliser of σ acts faithfully on a finite set. The proof also shows that the stabiliser of a point $\sigma \in \mathrm{Stab}_{\dagger}(\mathcal{C})$ injects into $\mathrm{GL}(d; \mathbb{Z})$ via the action on the central charge, so the torsion orders of stabilisers are bounded only in terms of d.

To deduce that the quotient is an orbifold, it remains to see that the action of $\mathrm{Auteq}_{\dagger}(\mathcal{C})$ is properly discontinuous, so admits slices. Bridgeland [2007, Lemma 6.4] proves that if two stability conditions σ, τ have the same central charge, then they co-incide or are at distance $d(\sigma, \tau) \geqslant 1$. Therefore, if the action was not properly discontinuous, there would be a stability condition σ and an infinite sequence of elements in $\mathrm{GL}(d, \mathbb{Z})$ which failed to displace a small ball around Z_σ. This easily yields a contradiction. Since autoequivalences act on $\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}), \mathbb{C})$ via the complexifications of integral linear maps, they preserve the real and imaginary subbundles of the tangent bundle. Thus, $\mathrm{Stab}_{\dagger}(\mathcal{C})$ has holomorphically trivial tangent bundle, and the tangent bundle of $\mathrm{Stab}_{\dagger}(\mathcal{C})/\mathrm{Auteq}_{\dagger}(\mathcal{C})$ is the complexification of a real bundle, so the odd Chern classes of the quotient orbifold are 2-torsion. □

The diagonal group $\mathrm{diag}(e^t, e^{-t}) \subset \mathrm{SL}(2, \mathbb{R})$ acts on $\mathrm{Stab}(\mathcal{C})$ by expanding and contracting the real and imaginary directions in $\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}), \mathbb{C})$, which are tangent to smooth Lagrangian subbundles of the tangent bundle (for the flat Kähler form on $\mathbb{C}^d$), somewhat reminiscent of an Anosov flow. If this flow on $\mathrm{Stab}(\mathcal{C})$ was the geodesic flow of a complete Riemannian metric, then classical results (non-existence of conjugate points) would imply that $\mathrm{Stab}(\mathcal{C})$ was a $K(\pi, 1)$. Despite the naivety of such reasoning, in known cases, the connected components of $\mathrm{Stab}(\mathcal{C})$ are contractible, or diffeomorphic to spaces independently conjectured to be contractible in the literature Allcock [2013], Kontsevich and Zorich [1997], and Qiu and Woolf [2014].

Corollary 3.4. *If* $\mathrm{Stab}_{\dagger}(\mathcal{C})$ *is contractible,* $\mathrm{Auteq}_{\dagger}(\mathcal{C})$ *has finite rational cohomological dimension.*

Proof. This is true for any discrete group Γ acting on a contractible finite-dimensional manifold Q with finite stabilisers, provided the action admits slices. (To prove the result, apply the Leray-Hirsch theorem to the projection $(Q \times E\Gamma) \to Q \times_{\Gamma} E\Gamma$, and note that the stalks of the pushforward of the constant sheaf $\mathbb{Q}$ vanish.) The existence of slices for the action follows from proper discontinuity. □

If a finitely generated countable group G has an embedding $G \times G \hookrightarrow G$ (e.g. the finitely presented "Thompson's group F") then it has infinite rational cohomological dimension Gandini [2012], so the contractibility hypothesis already excludes some "reasonable" countable groups from being autoequivalence groups.

It is tempting to believe contractibility holds, when it does, for some intrinsic geometric reason, e.g. that the canonical metric (3-1) is non-positively curved, making $\mathrm{Stab}(\mathcal{C})$ a complete CAT(0) space. (If the phases of σ-semistable objects are dense in S^1, then the $\mathrm{GL}^+(2,\mathbb{R})$-orbit of σ is free, and the metric on the quotient $\mathrm{GL}^+(2,\mathbb{R})/\mathbb{C}$ is the standard hyperbolic metric on the upper half-plane up to scale Woolf [2012].) This would be of dynamical relevance. Groups acting on weakly hyperbolic spaces with rank one elements (which typically exist) admit infinite-dimensional families of quasimorphisms, for instance the "counting" quasimorphisms of Epstein and Fujiwara [1997], see also Calegari [2009]. On the other hand, Malyutin [2011] shows that if $\Phi : G \to \mathbb{R}^3$ is defined by a triple of unbounded quasimorphisms, and if $S \subset G$ has bounded Φ-image, then S is transient, meaning that a random walk on G (defined with respect to any non-degenerate probability measure, i.e. one whose support generates G as a semigroup) will visit S only finitely many times almost surely. A prototypical example is that the reducible or periodic surface diffeomorphisms – those for which $\iota(\alpha, \phi^n(\beta))$ is not exponential for all α, β in (Section 2.1, (3)) – are transient in Γ_g, and random mapping classes of surfaces are pseudo-Anosov.

Corollary 3.5. *If* $\mathrm{Stab}_\dagger(\mathcal{C})$ *is complete CAT(0) and* $\mathrm{Auteq}_\dagger(\mathcal{C})$ *acts with rank one elements, then scl of the kth element of a random walk tends to infinity as k tends to infinity almost surely.*

One can analogously look for quantitative "unbounded generation" results. Borel and Harish-Chandra [1962] proved the isometry group of an arithmetic lattice acts on elements of fixed square with finitely many orbits. Suppose this homological statement lifts, and that $\mathcal{C}$ has only finitely many conjugacy classes of spherical object under autoequivalence. Quasimorphisms can be averaged to be homogeneous, hence constant on conjugacy classes, so then all spherical twists have uniformly bounded image under any Φ as above. If $S \subset G$ is transient and $N \subset G$ is finite, then $(S \cup S^{-1} \cup N)^k$ is transient for any fixed k. For $\mathcal{C} = \mathcal{F}(X)$ we have noted that one hopes a Lagrangian sphere defines an end to (a component meeting $\mathfrak{M}_X(J,\Omega)$ of) $\mathrm{Stab}(\mathcal{F}(X))$ corresponding to a nodal degeneration. Suppose that there is a partial compactification of $\mathcal{Q} = \mathrm{Stab}_\dagger(\mathcal{F}(X))/\mathrm{Auteq}_\dagger(X)$, or of $\mathfrak{M}_X(J,\Omega)$, with an irreducible primitive analytic divisor D parametrizing nodal degenerations of X. A choice of a closed oriented surface S transverse to D at one point exhibits, via the usual presentation for $\pi_1(S)$, the corresponding spherical twist as a product of commutators, and hence all conjugate spherical twists as products of the same number of commutators. Then Corollary 3.5 would imply that, for any fixed k, a random element of $\mathrm{Auteq}_{\dagger}(\mathcal{F}(X))$ would almost surely not be a product of fewer than k spherical twists.

This raises the question: how does one find *Floer-theoretic* conditions for sets of symplectomorphisms to lie in a bounded set under the image of a quasimorphism on $\pi_0\mathrm{Symp}(X)$?

3.3 Classification of objects. Consider some concrete computations. Since the space of stability conditions is expected to be contractible, the object of interest is the orbifold covering map

$$\pi : \mathrm{Stab}_{\dagger}(\mathcal{C}) \to \mathrm{Stab}_{\dagger}(\mathcal{C})/\mathrm{Auteq}_{\dagger}(\mathcal{C}) = \mathcal{Q}. \tag{3-2}$$

- (X, ω) is $(T^2, \omega_{\mathrm{st}})$. Then $\mathcal{Q}$ is a $\mathbb{C}^*$-bundle over $\mathfrak{h}/SL(2,\mathbb{Z})$, and $\mathrm{Auteq}(\mathcal{F}(X))$ is an extension of $SL(2;\mathbb{Z})$ by $\mathbb{Z} \oplus X \times X^{\vee}$ (acting by shift, translation and tensoring by flat bundles); Bridgeland [2007] and Polishchuk and Zaslow [1998].

- (X, ω) is $\{x^2 + y^2 + z^k = 1\} \subset (\mathbb{C}^3, \omega_{\mathbb{C}^3})$ and we consider the compact Fukaya category. Then $\mathcal{Q} = \mathrm{Conf}_k(\mathbb{C})$ parametrises polynomials with k distinct roots, cf. the monodromy discussion of (2-3); $\mathrm{Auteq}(\mathcal{F}(X))/[2] = Br_k$ is the braid group Thomas [2006] and Ishii, Ueda, and Uehara [2010].

- (X, ω) is (Σ, M), a surface with non-empty boundary containing a non-empty set of boundary marked points $M \subset \partial\Sigma$, and we consider the partially wrapped[1] Fukaya category $\mathcal{F}(\Sigma; M)$ stopped at M Sylvan [2015], whose objects include arcs ending on $\partial\Sigma \backslash M$. Then $\mathcal{Q}$ is a space of marked flat structures (meromorphic quadratic differentials) on Σ, up to diffeomorphism Haiden, Katzarkov, and Kontsevich [2017].

The computation of $\mathrm{Stab}(\mathcal{C})$, when feasible, often goes hand-in-hand with a classification for some class of objects of the category $\mathcal{C}$, which is of independent interest. (Whilst the objects of $\mathcal{F}(X)$ are *a priori* geometric, the objects of $\mathcal{F}(X)^{perf}$ are not, and a concrete interpretation of an arbitrary perfect module is rarely available.) In the cases above, one uses:

- Atiyah's classification of bundles on elliptic curves, which implies that a twisted complex on objects in $\mathcal{F}(T^2)$ is again quasi-represented by a simple closed curve with local system;

- Ishii-Uehara's proof Ishii and Uehara [2005] that all spherical objects in $\mathrm{Tw}\,\mathcal{F}(X_p)$ are in the braid group orbit of a fixed Lagrangian sphere;

[1] The compact Fukaya category $\mathcal{F}$ is proper but not smooth; the wrapped category $\mathcal{W}$ is smooth but not proper; the partially wrapped category, which depends on additional data – here, the set M – has both properties. It determines $\mathcal{W}$ as a localization, and (conjecturally, in nice cases) $\mathcal{F}$ as a subcategory of compact objects of $\mathcal{W}$.

- Haiden-Katzarkov-Kontsevich's proof that every indecomposable twisted complex of objects of $\mathcal{F}(\Sigma; M)$ is represented by an immersed curve with local system.

By appealing to either sheaf-theoretic properties of wrapped categories Lee [2015] or equivariant arguments Seidel [2012], one can deduce from the last example above that on a closed surface Σ_g, every spherical object is geometric, i.e. quasi-isomorphic to an immersed curve with local system. It is tempting to speculate that, starting from that point, one can recover the complex of curves and hence the classical mapping class group from the derived Fukaya category $D^\pi \mathcal{F}(\Sigma_g)$.

3.4 Measure and growth. Recall that $\mathrm{Stab}_\dagger(\mathcal{C})$ has a canonical measure $d\nu$ inherited from Lebesgue measure in $\mathrm{Hom}_{\mathbb{Z}}(K(\mathcal{C}), \mathbb{C})$. Since autoequivalences act linearly and integrally on $K(\mathcal{C})$, the measure is $\mathrm{Auteq}_\dagger(\mathcal{C})$-invariant, so descends to the quotient orbifold. This will essentially always have infinite measure, since there is a free $\mathbb{R}$-action on $\mathrm{Stab}(\mathcal{C})$, rescaling the central charge.

Hypothesis 3.6. *There is a submanifold* $\mathrm{Stab}_{=1,\dagger}(\mathcal{C})$ *which is a slice to the* $\mathbb{R}$*-action, which is invariant under the universal cover of* $SL(2,\mathbb{R})$*, and for which*

$$\mathrm{Stab}_{=1,\dagger}(\mathcal{C})/\mathrm{Auteq}_\dagger(\mathcal{C})$$

has finite measure.

Existence of a slice is usually not difficult, but the finite measure hypothesis is probably much more severe. Let S be a Riemann surface and $\phi \in H^0(2K_S)$ a quadratic differential with distinct zeroes. There is a local CY Kähler threefold

$$Y_\phi = \{(q_1, q_2, q_3) \in K_S \oplus K_S \oplus K_S \mid q_1^2 + q_2^2 + q_3^2 = \phi\} \tag{3-3}$$

which depends up to deformation only on the genus $g(S)$. The space of stability conditions modulo autoequivalences on $\mathcal{F}(Y_\phi)$ is conjecturally the moduli space $\mathcal{M}(S,\phi)$ of quadratic differentials with simple zeroes, and the normalisation $\int_S \phi \wedge \bar{\phi} = 1$ to surfaces of flat area 1 defines a slice to the $\mathbb{R}$-action. This has finite measure by deep results of Veech [1993]. For a K3 surface X°, $Z(E) = (\Omega, ch(E))$ for a period vector $\Omega \in \mathbb{N}(X^\circ) \otimes \mathbb{C}$, where $\mathbb{N}(X^\circ)$ is the extended Neron-Severi group $H^0 \oplus \mathrm{Pic} \oplus H^4$, and the condition $(\Omega, \bar{\Omega}) = 1$ defines a slice.

Given Hypothesis 3.6, one can play the following game (a well-known trope in the flat surfaces community; we follow Zorich [2006]). Fix a stability condition σ; the support property means that the central charges $Z(E)$ of σ-semistable objects are discrete in $\mathbb{C}$. For a compactly supported function $f \in C_{ct}(\mathbb{R}^2)$ we can then define $\hat{f}(\sigma) = \sum_{Z(E)} f(Z(E))$, summing over central charges represented by semistable E. A stronger version of Hypothesis 3.6 asks that the function $\sigma \mapsto \hat{f}(\sigma)$ be $d\nu$-integrable, yielding a linear function

$$f \mapsto \int_{\mathrm{Stab}_{=1,\dagger}(\mathcal{C})/\mathrm{Auteq}_{\dagger}(\mathcal{C})} \hat{f}\, d\nu$$

which is $SL_2(\mathbb{R})$-invariant. The only such functionals on $C_{ct}(\mathbb{R}^2)$ are given by the total area and the value at zero; the latter is irrelevant here since $Z(E) \neq 0$ for any semistable E, so one finds that

$$\text{For every } f \in C_{ct}(\mathbb{R}^2), \quad \int_{\mathrm{Stab}_{=1,\dagger}(\mathcal{C})/\mathrm{Auteq}_{\dagger}(\mathcal{C})} \hat{f}\, d\nu = \tau(\mathcal{C}) \int_{\mathbb{R}^2} f\, dxdy$$

where $\tau(\mathcal{C})$ is an invariant of the category and the choice of component / slice. Taking f to be the indicator function of a disc of increasing radius R, one sees that $\tau(\mathcal{C})$ measures the growth rate of K-theory classes represented by semi-stable objects as their mass increases. For $\mathcal{M}(S,\phi)$ above, $\tau(\mathcal{C})$ is a Siegel-Veech number, and controls the quadratic growth rate of special Lagrangian 3-spheres by volume in the 3-fold (3-3) (in that case, the special Lagrangian 3-sphere representative of the homology class is expected to be unique).

4 Three more extended examples

4.1 K3 surfaces. Let Y be an algebraic K3 surface of Picard rank $\rho \geqslant 1$. Bridgeland [2008] has given a conjectural description of a distinguished component of $\mathrm{Stab}(D^b(Y))$ as the universal cover of a period domain. Recall $\mathbb{N}(Y) = H^0(Y) \oplus \mathrm{Pic}(Y) \oplus H^4(Y)$, which carries the Mukai pairing $\langle (a,b,c),(a',b',c')\rangle = ac' + ca' - b\cdot b'$. Let

$$\Omega = \{u \in \mathbb{N}(Y)\otimes\mathbb{C} \mid \mathrm{Re}(u), \mathrm{Im}(u) \text{ span a positive definite } 2\text{–plane}\}.$$

Fix one of the two connected components of Ω, and remove the locally finite union of hyperplanes $\delta^{\perp}$ indexed by -2-vectors $\delta \in \mathbb{N}(Y)$ (with its Mukai form), to obtain a period domain $\mathcal{P}_0^+(Y)$. Then Bridgeland [ibid.] constructs a connected component $\mathrm{Stab}_{\dagger} \subset \mathrm{Stab}(D^b(Y))$ and proves that there is a Galois covering

$$\mathrm{Stab}_{\dagger} \to \mathcal{P}_0^+(Y)$$

which identifies the Torelli autoequivalence group with the group of deck transformations. Bridgeland conjectures that $\mathrm{Stab}_{\dagger}$ is simply-connected and Auteq-invariant, which would yield the exact sequence

$$1 \to \pi_1(\mathcal{P}_0^+(Y)) \to \mathrm{Auteq}_{CY}(D^b(Y)) \to \mathrm{Aut}^+(H^*(Y)) \to 1$$

where the Calabi-Yau autoequivalences Auteq_{CY} are those acting trivially on the Hochschild homology group $HH_2(Y)$, and the final term is the index two subgroup of the full Hodge isometry group of automorphisms preserving the orientation of a maximal positive definite subspace. The generators of the first term would be mapped to squared spherical twists. Separately, a conjecture of Allcock [2013] on Coxeter arrangements would imply that, if Bridgeland's conjecture holds, then $\mathrm{Stab}_{\dagger}(D^b(Y))$ is indeed a CAT(0) space, in particular contractible. When $\rho(Y) = 1$, both conjectures are known, which yields a complete determination of both that component of the space of stability conditions and the autoequivalence group.

Proposition 4.1. *Let (X, ω) be a Kähler K3 surface which is homologically mirror to an algebraic K3 surface Y of Picard rank one. (Minimality of the Picard rank on the mirror corresponds to an irrationality hypothesis on the Kähler form ω on X.) Let $G\mathcal{F}(X) = \mathrm{Auteq}(D^\pi \mathcal{F}(X, \omega))$.*

1. *$G\mathcal{F}(X)$ is finitely presented.*
2. *The Torelli subgroup is infinitely generated and torsion-free.*
3. *$G\mathcal{F}(X)$ contains no divisible elements.*
4. *$G\mathcal{F}(X)$ cannot obey a law.*

Sketch comments. The first three statements are essentially immediate from Bayer and Bridgeland [2017]; in fact the categorical Torelli group in the (mirror to a) Picard rank one case is a countably generated free group. scl for free products of cyclic groups was computed in Walker [2013], e.g. there are examples for which $G\mathcal{F}(X) = \mathbb{Z}/p * \mathbb{Z}$, and for $\mathbb{Z}/p * \mathbb{Z}$ with generators a, b of the factors, $scl([a, b]) = 1/2 - 1/p$; as remarked previously, non-vanishing of scl rules out laws. See Sheridan and Smith [2017b] for related ideas and details. □

Proofs of the Bridgeland and Allcock conjectures would eliminate the Picard rank one hypothesis. Results for $G\mathcal{F}(X)$ do not immediately yield results for the symplectic mapping class group $\pi_0\mathrm{Symp}(X, \omega)$, but these can sometimes be extracted with more care. Let (X, ω) be the mirror quartic, a crepant resolution of $\{\sum_{j=0}^3 x_j^4 + \lambda \prod_j x_j = 0\}/\Gamma$ for $\Gamma \cong (\mathbb{Z}/4)^{\oplus 2}$, equipped with an "irrational" toric Kähler form (one induced from its embedding into a toric resolution of $[\mathbb{P}^3/\Gamma]$ such that the areas of the resolution curves and of a hyperplane section are rationally independent). Then (see Sheridan and Smith [ibid.]):

Theorem 4.2 (Sheridan, Smith)**.** *For the mirror quartic with an irrational toric Kähler form, the group $\ker(\pi_0\mathrm{Symp}(X) \to \pi_0\mathrm{Diff}(X))$ is infinitely generated.*

The map π of (3-2) is (modulo $GL(2,\mathbb{R})$) the universal cover of the orbifold $\mathfrak{h}/\Gamma_0(2)^+$, where $\Gamma_0(2)^+ = \mathbb{Z}/2 * \mathbb{Z}/4$, and $\mathrm{Auteq}_{CY}(\mathcal{F}(X))/[2] = \mathbb{Z} * \mathbb{Z}/4$ where again the CY autoequivalencs are those acting trivially on $HH_2(\mathcal{F}(X))$, cf. Bayer and Bridgeland [2017] and Sheridan and Smith [2017b]. One infers the Dehn twist τ_L in a Lagrangian sphere $L \subset X$ admits no non-trivial root in $\pi_0\mathrm{Symp}(X)$ (being a smooth involution, it is its own $(2p+1)$-st root in $\pi_0\mathrm{Diff}(X)$, for any $p \geqslant 1$); $\pi_0\mathrm{Symp}(X)$ is not generated by torsion elements (which all map to the same factor in the abelianization of $\mathrm{Auteq}(\mathcal{F}(X))$), in contrast to Γ_g, etc.

Remark 4.3. *It is interesting to compare the proof in Bayer and Bridgeland [2017] with the cartoon description of stability conditions in terms of special Lagrangians given at the end of Section 3.1. For a Picard rank one K3 surface X°, the central charge is given by $Z(E) = \langle \Omega, ch(E)\rangle$ for a vector $\Omega \in \mathbb{N}(X^\circ) \otimes \mathbb{C} \cong \mathbb{C}^3$ whose real and imaginary parts span a positive definite two-plane. There is a unique negative definite vector $\theta \in \mathbb{N}(X^\circ) \otimes \mathbb{R}$ orthogonal to $\{Re(\Omega), Im(\Omega)\}$. Fix $E = \mathcal{O}_x$ the skyscraper sheaf of a point, and a stability condition σ for which the largest and smallest semistable factors $A_\pm$ of E have phases $\phi_+ > \phi_-$ respectively. Then Bayer and Bridgeland [ibid.] studies the flow on* $\mathrm{Stab}(X^\circ)$ *defined by*

$$d\Omega/dt = \xi.\theta \qquad \xi = i\,\exp(i\pi/2(\phi_+ + \phi_-))$$

which locally pushes $A_\pm$ towards one another, decreasing $\phi_+ - \phi_-$. They prove such flows can be patched together and eventually contract $\mathrm{Stab}(X^\circ)$ *to the geometric chamber where E is semistable. Translating back to the A-side, the cartoon is now that instead of mean curvature flow, one fixes the Lagrangian torus, and flows in the space of holomorphic volume forms to try to make it special.*

4.2 Quiver threefolds. Let (Q,W) be a quiver with potential. This determines a 3-dimensional Calabi-Yau category $\mathcal{C}(Q,W)$ Ginzburg [2006]. If (Q,W) has no loops, there is a "mutation" operation which yields another (Q',W') and a (pair of) derived equivalence(s) $\mathcal{C}(Q,W) \simeq \mathcal{C}(Q',W')$. In a number of interesting cases, these categories are related to Fukaya categories of threefolds:

1. The zero-potential on the two-cycle quiver (arrows labelled e, f) is realised within the compact Fukaya category of the affine quartic $\{x^2+y^2+(zt)^2=1\} \subset \mathbb{C}^4$, which is a plumbing of two 3-spheres along a circle (plumbed so the Lagrange surgery is an $S^1 \times S^2$), see Evans, Smith, and Wemyss [n.d.].

2. The potential $(ef)^2$ on the same quiver is realised by the compact Fukaya category of the complement of a smooth hyperplane section in the variety of complete flags

in $\mathbb{C}^3$, which is again a plumbing of two 3-spheres along a circle (plumbed so the Lagrange surgery is an S^3); potentials $(ef)^p$ on the two-cycle quiver, for prime $p > 2$, may arise in characteristic p from the corresponding plumbings where the surgery is a Lens space $L(p,1)$, see Evans, Smith, and Wemyss [ibid.].

3. The potential associated Labardini-Fragoso [2009] to an ideal triangulation of a marked bordered surface (S, M) is realised by the Fukaya category of a threefold which is a conic fibration over S with special fibres at M Smith [2015] (this is a cousin of the space from (3-3), where $M = \varnothing$).

The category $\mathcal{C}(Q, W)$ has a distinguished heart, equivalent to the category of nilpotent representations of the Jacobi algebra $\mathrm{Jac}(Q, W)$, with d simple objects up to isomorphism if Q has d vertices. This implies that a large subset $\mathcal{U} \subset \mathrm{Stab}(\mathcal{C}(Q, W))$ with non-empty open interior is a union of cells $\overline{\mathfrak{h}}^d$ (where $\overline{\mathfrak{h}}$ is the union of the upper half-plane and the negative real axis excluding zero), indexed by t-structures having hearts with finite length, glued together along their boundaries by the combinatorics of tilting (quiver mutation). This provides one of the most direct routes to (partial) computations of spaces of stability conditions. Often, the image of $\mathcal{U}$ under the natural circle action on $\mathrm{Stab}(\mathcal{C})/\langle[2]\rangle$ covers a path-component.

Let ϕ be a meromorphic quadratic differential on a surface S with at least one pole of order $\geqslant 2$, with p double poles, and distinct zeroes. Let $M \subset S$ be the set of poles. There is a threefold $Y_\phi \to S$, a variant of that from (3-3), now with empty fibres over poles of order > 2 and reducible fibres (singular at infinity) over double poles; a choice of component of each reducible fibre defines a class $\eta \in H^2(Y_\phi; \mathbb{Z}/2)$. Let $\mathcal{F}(Y_\phi; \eta)$ denote the subcategory of the η-sign-twisted Fukaya category split-generated by Lagrangian spheres. Then (see Bridgeland and Smith [2015] and Smith [2015]):

Theorem 4.4 (Bridgeland, Smith)**.** *There is an equivalence $\mathcal{F}(Y_\phi; \eta) \simeq \mathcal{C}(Q, W)$ for (Q, W) the quiver with potential associated to any ideal triangulation of (S, M). Moreover,*

$$1 \to \mathrm{Sph}(S, M) \to \mathrm{Auteq}(\mathcal{C}(Q, W)) \to \Gamma^{\pm}(S, M) \to 1 \tag{4-1}$$

where $\Gamma^{\pm}$ is an extension of the mapping class group $\Gamma(S, M)$ by $(\mathbb{Z}/2)^p$.

The first factor in (4-1) acts through spherical twists and admits a natural representation to $\pi_0\mathrm{Symp}_{ct}(Y_\phi)$ whilst the quotient factor acts via non-compactly-supported elements of $\pi_0\mathrm{Symp}(Y_\phi)$. In particular, the natural map $\mathrm{Sph}(S, M) \to \pi_0\mathrm{Symp}_{ct}(Y_\phi)$ is actually split. Stable objects in Y_ϕ are all given by special Lagrangian 3-spheres or $S^1 \times S^2$'s, corresponding to open and closed saddle connections in the flat metric on (S, ϕ).

There are 3-folds $Y_{\phi,\psi}$ associated to a pair $(\phi,\psi) \in H^0(K_S^{\oplus 2}) \oplus H^0(K_S^{\oplus 3})$, fibred over S by A_2-Milnor fibres rather than A_1-Milnor fibres: given line bundles L_1, L_2 over S with $K_S = L_1 L_2$,

$$Y_{\phi,\psi} = \{(x,y,z) \in L_1^3 \oplus L_2^3 \oplus L_1 L_2 \,|\, xy = z^3 + \phi \cdot z + \psi\}.$$

In this case the conjectural embedding of a moduli space of pairs (ϕ,ψ) into $\mathrm{Stab}(\mathcal{F}(Y_{\phi,\psi}))$ cannot be onto an open set for dimension reasons (cf. Section 3.1). The DT-counting invariants for semistable objects in K-theory class $d\gamma$ can have exponential growth in d Galakhov, Longhi, Mainiero, Moore, and Neitzke [2013] and have interesting algebraic generating functions Mainiero [2016]. The 3-fold $Y_{\phi,\psi}$ now contains a special Lagrangian submanifold $L \cong (S^1 \times S^2)\#(S^1 \times S^2)$, obtained from surgery of two 3-spheres lying over tripods which meet at three end-points, and the wild representation theory of $\pi_1(L)$ may be responsible for the exponential growth of stable objects (flat bundles over L) on the symplectic side. It would be interesting to know if (3.6) is related to polynomial growth of DT-invariants.

4.3 Cubic four-folds. The derived category of a cubic four-fold $Y \subset \mathbb{P}^5$ admits a semi-orthogonal decomposition, the interesting piece of which is a CY2-category $\mathcal{A}_Y$ introduced by Kuznetsov [2010]. These categories are of symplectic nature, cf. Huybrechts [2017, Proposition 2.17] and Sheridan and Smith [2017a]. Let E be the (Fermat) elliptic curve with a non-trivial $\mathbb{Z}/3$-action, generated by ξ, and X be the K3 surface which is the crepant resolution of $(E \times E)/\langle(\xi,\xi^{-1})\rangle$. (This is sometimes called the "most algebraic" K3 surface; it has Picard rank 20 and, amongst such K3's, has smallest possible discriminant.) Then for certain toric Kähler forms ω on X which are "irrational" (again meaning the areas of the resolution curves and a hyperplane section are linearly independent over $\mathbb{Q}$), there is an equivalence Sheridan [2017] and Sheridan and Smith [2017a] (strictly, this requires incorporating certain immersed Lagrangian tori into $\mathcal{F}(X,\omega)$)

$$D^\pi \mathcal{F}(X,\omega) \simeq \mathcal{A}_{Y_{d(\omega)}} \subset D^b(Y_{d(\omega)})$$

where the valuations of the coefficients in the equation defining the cubic Y over Λ are determined by the choice of Kähler form. For sufficiently general Y, the space of stability conditions has been computed by Bayer, Lahoz, Macrì, and Stellari [2017], following a direct computation of autoequivalences due to Huybrechts [2017], and one finds (cf. Sheridan and Smith [2017a, n.d.]):

Theorem 4.5 (Sheridan, Smith)**.** *For the most algebraic K3 surface X with an irrational toric Kähler form, the map $\rho : \pi_0\mathrm{Symp}(X,\omega) \longrightarrow \mathrm{Auteq}(\mathcal{F}(X,\omega))/[2]$ has image $\mathbb{Z}/3$. If $Z(X,\omega) = \ker(\rho)$ then $\pi_0\mathrm{Symp}(X,\omega) = Z(X,\omega) \rtimes \mathbb{Z}/3$.*

The image of ρ is generated by the obvious residual diagonal action on $E \times E$, which lifts to X. Theorem 4.5 reduces the task of understanding the symplectic mapping class group of (X, ω) to that of understanding what Floer theory doesn't see, which is about as far as we could hope to come (one can draw similar conclusions in the setting of Theorem 4.2, see Sheridan and Smith [2017b]). More concretely, for such irrational Kähler forms one learns that $HF(\phi)$ can take only one of three possibilties for any symplectomorphism ϕ, and any ϕ has trivial Floer-theoretic entropy, etc. From the Künneth theorem, this has consequences for fixed points of symplectomorphisms of $X \times T^2$, say, which – in the same vein as the Arnol'd conjecture – go beyond information extractable from smooth topology (stabilising by taking product with T^2 serves to kill information from the Lefschetz theorem).

It is interesting to imagine turning around the direction of mirror symmetry in this case. Naively, one would predict some relation between the categories

$$D^{\pi}\mathcal{F}(Y;0) \qquad \text{and} \qquad D^b(X) = D(E \times E)^{\mathbb{Z}/3} \tag{4-2}$$

where $\mathcal{F}(Y;0)$ is the nilpotent summand of the Fukaya category (corresponding in (2-1) to the zero-eigenvalue). It seems far from obvious that these should be equivalent, and the actual relation, if any, might be more subtle. Nonetheless, although $\mathcal{F}(Y)$ is not $\mathbb{Z}$-graded as Y is Fano, the summand $\mathcal{F}(Y;0)$ *is* expected to admit a $\mathbb{Z}$-grading, whence one could talk about stability conditions. Let $\mathcal{M}_{3,d}$ denote the moduli space of cubic d-folds. These are well-studied spaces (and are famously CAT(0) for $d = 2, 3$ Allcock, Carlson, and Toledo [2002, 2011]). Starting from the lattice-theoretic co-incidence $H^2(X;\mathbb{Z}) \supset \mathrm{Pic}(X)^{\perp} = -A_2 = \langle h^2\rangle^{\perp} \subset H^4(Y;\mathbb{Z})$, where $h^2 \in H^4(Y;\mathbb{Z})$ is the class defined by a hyperplane, results of Laza [2010] yield an embedding (also observed by R. Potter)

$$\mathcal{M}_{3,4} \longrightarrow \mathrm{Auteq}_{CY}(X)\backslash \mathrm{Stab}_{\dagger}(D^b(X))/\widetilde{GL}^{+}(2,\mathbb{R})$$

onto the complement of an explicit divisor Δ (associated to the locally finite hyperplane arrangement defined by classes of square -6, or equivalently the complement in the period domain of the divisors associated to classes of square -2 and -6). Bridgeland's conjecture and some concrete relation in (4-2) might then give insight into the monodromy homomorphism

$$\pi_1(\mathcal{M}_{3,4}) \longrightarrow \pi_0\mathrm{Symp}(Y) \longrightarrow \mathrm{Auteq}(D^{\pi}\mathcal{F}(Y;0)).$$

In this way, one could hope to use stability conditions on $K3$ surfaces to attack classical problems related to the symplectic monodromy of hypersurfaces.

4.4 Clusters. Semantic sensitivities notwithstanding, we end with a digression. It may be useful to point out where much of the activity in the subject is concentrated. If (Q, W)

satisfies suitable non-degeneracy assumptions, one can associate to (Q, W) two spaces: $\mathrm{Stab}(\mathcal{C}(Q, W))$ and the cluster variety $\mathcal{X}(Q, W)$. The first is glued together out of chambers $\overline{\mathfrak{h}}^d$ indexed by the vertices of the tilting tree, and the second is glued from birational maps of algebraic tori $(\mathbb{C}^*)^d$ indexed by the same data. It is believed that there is a (complicated, transcendental) complex Lagrangian submanifold $B \subset \mathrm{Stab}(\mathcal{C})$ and an algebraic integrable system (with compact complex torus fibres) $\mathcal{T} \to B$ for which $\mathcal{T}$ and $\mathcal{X}$ are diffeomorphic, naturally equipped with different complex structures belonging to a single hyperkähler family Gaiotto, Moore, and Neitzke [2013] and Neitzke [2014]. The explicit diffeomorphism should be obtained from a Riemann-Hilbert problem, whose definition and solution involves the moduli stacks of stable objects and their Donaldson-Thomas theory Bridgeland [2016].

A genus d Lagrangian surface $\Sigma_d \subset X^4$ in a symplectic four-manifold defines a chart $(\mathbb{C}^*)^d$ in a tentative mirror to X, and one can use the complexity of cluster atlases to prove existence theorems for infinite families of Lagrangian surfaces Shende, Treumann, and Williams [2016]. This is a striking connection back to symplectic topology, but one that it seems hard to formulate on the space of stability conditions directly.

References

Daniel Allcock (2013). "Completions, branched covers, Artin groups, and singularity theory". *Duke Math. J.* 162.14, pp. 2645–2689. MR: 3127810 (cit. on pp. 996, 1001).

Daniel Allcock, James A. Carlson, and Domingo Toledo (2002). "The complex hyperbolic geometry of the moduli space of cubic surfaces". *J. Algebraic Geom.* 11.4, pp. 659–724. MR: 1910264 (cit. on p. 1005).

– (2011). "The moduli space of cubic threefolds as a ball quotient". *Mem. Amer. Math. Soc.* 209.985, pp. xii+70. MR: 2789835 (cit. on p. 1005).

Gilbert Baumslag and Charles F. Miller III (1988). "Some odd finitely presented groups". *Bull. London Math. Soc.* 20.3, pp. 239–244. MR: 931184 (cit. on p. 989).

Arend Bayer and Tom Bridgeland (2017). "Derived automorphism groups of K3 surfaces of Picard rank 1". *Duke Math. J.* 166.1, pp. 75–124. MR: 3592689 (cit. on pp. 1001, 1002).

Arend Bayer, Martí Lahoz, Emanuele Macrì, and Paolo Stellari (Mar. 2017). "Stability conditions on Kuznetsov components". arXiv: 1703.10839 (cit. on p. 1004).

Arend Bayer and Emanuele Macrì (2011). "The space of stability conditions on the local projective plane". *Duke Math. J.* 160.2, pp. 263–322. MR: 2852118 (cit. on p. 994).

Armand Borel and Harish-Chandra (1962). "Arithmetic subgroups of algebraic groups". *Ann. of Math. (2)* 75, pp. 485–535. MR: 0147566 (cit. on p. 997).

Tom Bridgeland (2007). "Stability conditions on triangulated categories". *Ann. of Math. (2)* 166.2, pp. 317–345. MR: 2373143 (cit. on pp. 987, 994, 996, 998).

– (2008). "Stability conditions on $K3$ surfaces". *Duke Math. J.* 141.2, pp. 241–291. MR: 2376815 (cit. on p. 1000).

– (Nov. 2016). "Riemann-Hilbert problems from Donaldson-Thomas theory". arXiv: 1611.03697 (cit. on p. 1006).

Tom Bridgeland and Ivan Smith (2015). "Quadratic differentials as stability conditions". *Publ. Math. Inst. Hautes Études Sci.* 121, pp. 155–278. MR: 3349833 (cit. on pp. 987, 1003).

William Browder (1967). "Diffeomorphisms of 1-connected manifolds". *Trans. Amer. Math. Soc.* 128, pp. 155–163. MR: 0212816 (cit. on p. 989).

Danny Calegari (2009). *scl*. Vol. 20. MSJ Memoirs. Mathematical Society of Japan, Tokyo, pp. xii+209. MR: 2527432 (cit. on pp. 988, 997).

David B. A. Epstein and Koji Fujiwara (1997). "The second bounded cohomology of word-hyperbolic groups". *Topology* 36.6, pp. 1275–1289. MR: 1452851 (cit. on p. 997).

J. D. Evans, I. Smith, and M. Wemyss (n.d.). Manuscript in preparation (cit. on pp. 1002, 1003).

Benson Farb and Dan Margalit (2012). *A primer on mapping class groups*. Vol. 49. Princeton Mathematical Series. Princeton University Press, Princeton, NJ, pp. xiv+472. MR: 2850125 (cit. on p. 988).

Kenji Fukaya, Yong-Geun Oh, Hiroshi Ohta, and Kaoru Ono (2009). *Lagrangian intersection Floer theory: anomaly and obstruction. Part II*. Vol. 46. AMS/IP Studies in Advanced Mathematics. American Mathematical Society, Providence, RI; International Press, Somerville, MA, i–xii and 397–805. MR: 2548482 (cit. on p. 987).

Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke (2013). "Wall-crossing, Hitchin systems, and the WKB approximation". *Adv. Math.* 234, pp. 239–403. MR: 3003931 (cit. on p. 1006).

Dmitry Galakhov, Pietro Longhi, Tom Mainiero, Gregory W. Moore, and Andrew Neitzke (May 2013). "Wild Wall Crossing and BPS Giants". arXiv: 1305.5454 (cit. on p. 1004).

Søren Galatius and Oscar Randal-Williams (2017). "Homological stability for moduli spaces of high dimensional manifolds. II". *Ann. of Math. (2)* 186.1, pp. 127–204. MR: 3665002 (cit. on p. 992).

Giovanni Gandini (2012). "Cohomological invariants and the classifying space for proper actions". *Groups Geom. Dyn.* 6.4, pp. 659–675. MR: 2996406 (cit. on p. 997).

Victor Ginzburg (Dec. 2006). "Calabi-Yau algebras". arXiv: math/0612139 (cit. on p. 1002).

Sergey Gorchinskiy and Dmitri Orlov (2013). "Geometric phantom categories". *Publ. Math. Inst. Hautes Études Sci.* 117, pp. 329–349. MR: 3090263 (cit. on p. 994).

M. Gromov (1985). "Pseudo holomorphic curves in symplectic manifolds". *Invent. Math.* 82.2, pp. 307–347. MR: 809718 (cit. on p. 991).

F. Haiden, L. Katzarkov, and M. Kontsevich (2017). "Flat surfaces and stability structures". *Publ. Math. Inst. Hautes Études Sci.* 126, pp. 247–318. arXiv: 1409.8611. MR: 3735868 (cit. on p. 998).

A. E. Hatcher (1978). "Concordance spaces, higher simple-homotopy theory, and applications". In: *Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 1*. Proc. Sympos. Pure Math., XXXII. Amer. Math. Soc., Providence, R.I., pp. 3–21. MR: 520490 (cit. on p. 989).

Hans-Joachim Hein and Song Sun (2017). "Calabi-Yau manifolds with isolated conical singularities". *Publ. Math. Inst. Hautes Études Sci.* 126, pp. 73–130. arXiv: 1607.02940. MR: 3735865 (cit. on p. 995).

W. C. Hsiang and R. W. Sharpe (1976). "Parametrized surgery and isotopy". *Pacific J. Math.* 67.2, pp. 401–459. MR: 0494165 (cit. on p. 989).

D. Huybrechts (2012). "Stability conditions via spherical objects". *Math. Z.* 271.3-4, pp. 1253–1270. MR: 2945606 (cit. on p. 993).

Daniel Huybrechts (2016a). *Lectures on K3 surfaces*. Vol. 158. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, pp. xi+485. MR: 3586372 (cit. on p. 993).

– (2016b). "On derived categories of K3 surfaces, symplectic automorphisms and the Conway group". In: *Development of moduli theory—Kyoto 2013*. Vol. 69. Adv. Stud. Pure Math. Math. Soc. Japan, [Tokyo], pp. 387–405. MR: 3586514 (cit. on p. 995).

– (2017). "The K3 category of a cubic fourfold". *Compos. Math.* 153.3, pp. 586–620. MR: 3705236 (cit. on p. 1004).

Akira Ishii, Kazushi Ueda, and Hokuto Uehara (2010). "Stability conditions on A_n-singularities". *J. Differential Geom.* 84.1, pp. 87–126. MR: 2629510 (cit. on p. 998).

Akira Ishii and Hokuto Uehara (2005). "Autoequivalences of derived categories on the minimal resolutions of A_n-singularities on surfaces". *J. Differential Geom.* 71.3, pp. 385–435. MR: 2198807 (cit. on p. 998).

Nikolai V. Ivanov (1997). "Automorphism of complexes of curves and of Teichmüller spaces". *Internat. Math. Res. Notices* 14, pp. 651–666. MR: 1460387 (cit. on p. 989).

Dominic Joyce (2015). "Conjectures on Bridgeland stability for Fukaya categories of Calabi-Yau manifolds, special Lagrangians, and Lagrangian mean curvature flow". *EMS Surv. Math. Sci.* 2.1, pp. 1–62. MR: 3354954 (cit. on p. 995).

Tatsuji Kambayashi, Masayoshi Miyanishi, and Mitsuhiro Takeuchi (1974). *Unipotent algebraic groups*. Lecture Notes in Mathematics, Vol. 414. Springer-Verlag, Berlin-New York, pp. v+165. MR: 0376696 (cit. on p. 989).

Ailsa M. Keating (2014). "Dehn twists and free subgroups of symplectic mapping class groups". *J. Topol.* 7.2, pp. 436–474. MR: 3217627 (cit. on p. 992).

Mikhail Khovanov and Paul Seidel (2002). "Quivers, Floer cohomology, and braid group actions". *J. Amer. Math. Soc.* 15.1, pp. 203–271. MR: 1862802 (cit. on p. 992).

M. Kontsevich and A. Zorich (Jan. 1997). "Lyapunov exponents and Hodge theory". arXiv: hep-th/9701164 (cit. on p. 996).

Alexander Kuznetsov (2010). "Derived categories of cubic fourfolds". In: *Cohomological and geometric approaches to rationality problems*. Vol. 282. Progr. Math. Birkhäuser Boston, Inc., Boston, MA, pp. 219–243. MR: 2605171 (cit. on p. 1004).

Daniel Labardini-Fragoso (2009). "Quivers with potentials associated to triangulated surfaces". *Proc. Lond. Math. Soc. (3)* 98.3, pp. 797–839. MR: 2500873 (cit. on p. 1003).

Radu Laza (2010). "The moduli space of cubic fourfolds via the period map". *Ann. of Math. (2)* 172.1, pp. 673–711. MR: 2680429 (cit. on p. 1005).

Heather Ming Lee (2015). *Homological mirror symmetry for open Riemann surfaces from pair-of-pants decompositions*. Thesis (Ph.D.)–University of California, Berkeley. ProQuest LLC, Ann Arbor, MI, p. 51. arXiv: 1608.04473. MR: 3450266 (cit. on p. 999).

Tom Mainiero (June 2016). "Algebraicity and Asymptotics: An explosion of BPS indices from algebraic generating series". Preprint. arXiv: 1606.02693 (cit. on p. 1004).

A. V. Malyutin (2011). "Quasimorphisms, random walks, and transient subsets in countable groups". *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 390.Teoriya Predstavleniĭ, Dinamicheskie Sistemy, Kombinatornye Metody. XX, pp. 210–236, 309–310. MR: 2870236 (cit. on p. 997).

Howard A. Masur and Yair N. Minsky (1999). "Geometry of the complex of curves. I. Hyperbolicity". *Invent. Math.* 138.1, pp. 103–149. MR: 1714338 (cit. on p. 989).

Shigeyuki Morita (2007). "Introduction to mapping class groups of surfaces and related groups". In: *Handbook of Teichmüller theory. Vol. I*. Vol. 11. IRMA Lect. Math. Theor. Phys. Eur. Math. Soc., Zürich, pp. 353–386. MR: 2349674 (cit. on p. 988).

Andrew Neitzke (2014). "Notes on a new construction of hyperkahler metrics". In: *Homological mirror symmetry and tropical geometry*. Vol. 15. Lect. Notes Unione Mat. Ital. Springer, Cham, pp. 351–375. MR: 3330790 (cit. on p. 1006).

K. Ono (2006). "Floer-Novikov cohomology and the flux conjecture". *Geom. Funct. Anal.* 16.5, pp. 981–1020. MR: 2276532 (cit. on p. 990).

Alexander Polishchuk and Eric Zaslow (1998). "Categorical mirror symmetry: the elliptic curve". *Adv. Theor. Math. Phys.* 2.2, pp. 443–470. MR: 1633036 (cit. on p. 998).

Yu Qiu and Jon Woolf (July 2014). "Contractible stability spaces and faithful braid group actions". Preprint. arXiv: 1407.5986 (cit. on p. 996).

Randal-Williams (2015). "On diffeomorphisms acting on almost complex structures" (cit. on p. 991).

H. L. Royden (1970). "Report on the Teichmüller metric". *Proc. Nat. Acad. Sci. U.S.A.* 65, pp. 497–499. MR: 0259115 (cit. on p. 989).

Y. Ruan (1993). "Symplectic topology and extremal rays". *Geom. Funct. Anal.* 3.4, pp. 395–430. MR: 1223437 (cit. on p. 991).

P. Seidel (1997). "π_1 of symplectic automorphism groups and invertibles in quantum homology rings". *Geom. Funct. Anal.* 7.6, pp. 1046–1095. MR: 1487754 (cit. on p. 992).

Paul Seidel (1999). "Lagrangian two-spheres can be symplectically knotted". *J. Differential Geom.* 52.1, pp. 145–171. MR: 1743463 (cit. on p. 990).

– (2008). *Fukaya categories and Picard-Lefschetz theory*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, pp. viii+326. MR: 2441780 (cit. on p. 987).

– (2012). "Lagrangian homology spheres in (A_m) Milnor fibres via $\mathbb{C}^*$-equivariant A_∞-modules". *Geom. Topol.* 16.4, pp. 2343–2389. MR: 3033519 (cit. on p. 999).

Vivek Shende, David Treumann, and Harold Williams (Mar. 2016). "On the combinatorics of exact Lagrangian surfaces". Preprint. arXiv: 1603.07449 (cit. on p. 1006).

Nick Sheridan (Sept. 2017). "Versality of the relative Fukaya category". Preprint. arXiv: 1709.07874 (cit. on p. 1004).

Nick Sheridan and Ivan Smith (n.d.). Work in progress (cit. on p. 1004).

– (Sept. 2017a). "Homological mirror symmetry for generalized Greene-Plesser mirrors". Preprint. arXiv: 1709.08937 (cit. on p. 1004).

– (Sept. 2017b). "Symplectic topology of K3 surfaces via mirror symmetry". Preprint. arXiv: 1709.09439 (cit. on pp. 987, 993, 1001, 1002, 1005).

Ivan Smith (2015). "Quiver algebras as Fukaya categories". *Geom. Topol.* 19.5, pp. 2557–2617. MR: 3416110 (cit. on pp. 987, 1003).

Dennis Sullivan (1977). "Infinitesimal computations in topology". *Inst. Hautes Études Sci. Publ. Math.* 47, 269–331 (1978). MR: 0646078 (cit. on p. 989).

Zachary Aaron Sylvan (2015). *On partially wrapped Fukaya categories*. Thesis (Ph.D.)–University of California, Berkeley. ProQuest LLC, Ann Arbor, MI, p. 104. arXiv: 1604.02540. MR: 3427304 (cit. on p. 998).

R. P. Thomas (2006). "Stability conditions and the braid group". *Comm. Anal. Geom.* 14.1, pp. 135–161. MR: 2230573 (cit. on p. 998).

Orsola Tommasi (2014). "Stable cohomology of spaces of non-singular hypersurfaces". *Adv. Math.* 265, pp. 428–440. MR: 3255466 (cit. on p. 992).

Dmitry Tonkonog (2015). "Commuting symplectomorphisms and Dehn twists in divisors". *Geom. Topol.* 19.6, pp. 3345–3403. MR: 3447106 (cit. on p. 990).

William A. Veech (1993). "Flat surfaces". *Amer. J. Math.* 115.3, pp. 589–689. MR: 1221838 (cit. on p. 999).

Alden Walker (2013). "Stable commutator length in free products of cyclic groups". *Exp. Math.* 22.3, pp. 282–298. MR: 3171093 (cit. on p. 1001).

Chin-Lung Wang (1997). "On the incompleteness of the Weil-Petersson metric along degenerations of Calabi-Yau manifolds". *Math. Res. Lett.* 4.1, pp. 157–171. MR: 1432818 (cit. on p. 995).

J. Woolf (2012). "Some metric properties of spaces of stability conditions". *Bull. Lond. Math. Soc.* 44.6, pp. 1274–1284. MR: 3007660 (cit. on pp. 995, 997).

Anton Zorich (2006). "Flat surfaces". In: *Frontiers in number theory, physics, and geometry. I*. Springer, Berlin, pp. 437–583. MR: 2261104 (cit. on p. 999).

Received 2017-11-12.

Ivan Smith
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road CB3 0WB
U.K.
I.Smith@dpmms.cam.ac.uk
is200@cam.ac.uk

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1011–1030)

DEGENERATIONS AND MODULI SPACES IN KÄHLER GEOMETRY

Song Sun (孙崧)

Abstract

We report some recent progress on studying degenerations and moduli spaces of canonical metrics in Kähler geometry, and the connection with algebraic geometry, with a particular emphasis on the case of Kähler–Einstein metrics.

1 Introduction

One of the most intriguing features in Kähler geometry is the interaction between differential geometric and algebro-geometric aspects of the theory. In complex dimension one, the classical uniformization theorem provides a unique conformal metric of constant curvature -1 on any compact Riemann surface of genus bigger than 1. This analytic result turns out to be deeply connected to the algebraic fact that any such Riemann surface can be embedded in projective space as a *stable* algebraic curve in the sense of geometric invariant theory. It is also well-known that the moduli space of smooth algebraic curves of genus bigger than 1 can be compactified by adding certain singular curves with nodes, locally defined by the equation $xy = 0$, which gives rise to the *Deligne–Mumford compactification*. This is compatible with the differential geometric compactification using hyperbolic metrics, in the sense that whenever a node forms, locally the corresponding metric splits into the union of two hyperbolic cusps with infinite diameter. Intuitively, one can view the latter as a canonical differential geometric object associated to a nodal singularity.

On higher dimensional Kähler manifolds, a natural generalization of constant curvature metrics is the notion of a *canonical Kähler metric*, which is governed by some elliptic partial differential equation. In this article we will mainly focus on Calabi's *extremal Kähler metrics*. Let X be a compact Kähler manifold of dimension n, and let $\mathcal{H}$ be a

Partially supported by NSF grant DMS-1708420, the Alfred P. Sloan Fellowship, and a Simons Foundation grant (488633).

MSC2010: primary 53C55; secondary 53C21, 53C44, 14J10, 14J45.

Kähler class on X, i.e. the space of Kähler metrics on X in a given de Rham cohomology class (which we shall denote by $[\mathcal{H}] \in H^2(X;\mathbb{R})$). For example, any ample line bundle L gives rise to a Kähler class with $[\mathcal{H}] = 2\pi c_1(L)$. If we fix one Kähler metric $\omega \in \mathcal{H}$, then any other Kähler metric in $\mathcal{H}$ is of the form $\omega + i\partial\bar{\partial}\phi$ for some smooth function ϕ.

A Kähler metric ω is called *extremal* if it is the critical point of the Calabi functional $Ca(\omega) = \int_X S(\omega)^2\omega^n$ on $\mathcal{H}$, where $S(\omega)$ denotes the scalar curvature function of ω. Equivalently, by calculating the Euler–Lagrange equation, this means that the gradient vector field $\nabla_\omega S(\omega)$ is holomorphic. A special case is when the vector field vanishes, then ω is a *constant scalar curvature Kähler* metric. If moreover the first Chern class $c_1(X)$ is proportional to $[\mathcal{H}]$, say $2\pi c_1(X) = \lambda[\mathcal{H}]$, then this further reduces to the *Kähler–Einstein equation* $Ric(\omega) = \lambda\omega$.

There are several well-known fundamental questions that aim to build connections between the analytic theory of these metrics and algebraic geometry.

(1). **Existence**: When does $(X, \mathcal{H})$ contain an extremal Kähler metric?

To find an extremal Kähler metric amounts to solving a difficult non-linear elliptic PDE. The famous *Yau–Tian–Donaldson conjecture* states that the solvability of this PDE is equivalent to *K-stability* of $(X, \mathcal{H})$. K-stability is a complex/algebro-geometric notion; roughly speaking, it is tested by the positivity of certain numerical invariant, the *Donaldson–Futaki invariant*, associated to $\mathbb{C}^*$ equivariant *flat* degenerations of $(X, \mathcal{H})$. It is analogous to the Hilbert–Mumford criterion for stability in geometric invariant theory.

The Yau–Tian–Donaldson conjecture extends the *Calabi conjecture* on the existence of Kähler–Einstein metrics. In this special case we have a positive answer.

Theorem 1.1. *Let X be a compact Kähler manifold and $\mathcal{H}$ be a Kähler class with $2\pi c_1(X) = \lambda[\mathcal{H}]$.*

- *(Yau [1978]) If $\lambda = 0$, then there is a unique Kähler metric $\omega \in \mathcal{H}$ with $Ric(\omega) = 0$. This ω is usually referred to as a Calabi–Yau metric.*
- *(Yau [ibid.], Aubin [1976]) If $\lambda < 0$ (in which case X is of general type), then there is a unique Kähler metric $\omega \in \mathcal{H}$ with $Ric(\omega) = -\lambda\omega$.*
- *If $\lambda > 0$ (in which case X is a Fano manifold), then there is a Kähler metric $\omega \in \mathcal{H}$ with $Ric(\omega) = \lambda\omega$ if and only if X is K-stable. The "only if" direction is due to various authors in different generality including Tian [1997], Stoppa [2009], and R. J. Berman [2016], and the "if" direction is due to Chen–Donaldson–Sun (c.f. Chen, Donaldson, and Sun [2015a,b,c]), and later other proofs can be found in Datar and Székelyhidi [2016], Chen, Sun, and B. Wang [2015], and R. Berman, Boucksom, and Jonsson [2015].*

(2). **Compactification of moduli space and singularities**

It is known (c.f. Chen and Tian [2008] and R. J. Berman and Berndtsson [2017]) that extremal metric in a Kähler class $\mathcal{K}$, if exists, is unique up to holomorphic transformations of X, so it canonically represents the complex geometry of $(X, \mathcal{K})$. Locally one can deform either the Kähler class $\mathcal{K}$ or the complex structure on X, and the corresponding deformation theory of extremal Kähler metrics has been well-studied (c.f. LeBrun and Simanca [1994]). Globally to compactify the moduli space one needs to know how to take limits and what is the structure of the limits, especially near the singularities. This has interesting connection with the compactification from the algebro-geometric viewpoint, and is also intimately related to the existence question in (1).

(3). **Optimal degenerations**

There are two important geometric evolution equations, namely, the *Kähler–Ricci flow*, and the *Calabi flow*, that try to evolve an arbitrary Kähler metric towards a canonical metric. When $(X, \mathcal{K})$ is K-stable, then one expects the flow to exist for all time and converge to a canonical metric; when $(X, \mathcal{K})$ is not K-stable, then one expects the flow to generate a degeneration to some canonical geometric object associated to $(X, \mathcal{K})$, which is *optimal* in suitable sense.

We refer the readers to the article Donaldson [2018] in this proceeding for an overview on the existence question, and related discussion on K-stability. In this article we will report progress towards (2) and (3), mostly focusing on the case of Kähler–Einstein metrics.

Acknowledgments. I would like to thank Xiuxiong Chen, Simon Donaldson, Weiyong He, Hans-Joachim Hein, Chi Li, Yuji Odaka, Cristiano Spotti, Jingzhou Sun, Bing Wang, Yuanqi Wang, and Chengjian Yao for various fruitful discussions and collaborations on Kähler geometry. I am also grateful to Hans-Joachim Hein, Yuji Odaka, and Cristiano Spotti for helpful comments.

2 Gromov–Hausdorff limits

We consider a sequence of compact Kähler manifolds (X_i, ω_i) of dimension n. Throughout this section we will impose the following hypothesis

- $[\omega_i] = 2\pi c_1(L_i)$ for some ample line bundle L_i over X_i.

- $|Ric(\omega_i)| \leq \Lambda$ for some fixed $\Lambda > 0$.
- The diameter of (X_i, ω_i) is uniformly bounded above by $D > 0$.

These ensure the sequence of metrics to satisfy a *volume non-collapsing condition*, i.e. there exists $\kappa > 0$ such that for all i and all $p \in X_i$, $Vol(B(p,r)) \geq \kappa r^{2n}$ for all $r \in (0,1]$.

By the convergence theory of Riemannian manifolds one can pass to a subsequence and extract a *Gromov–Hausdorff limit* X_∞, which is a compact length space. This seems to be a very rough process at first sight but the important fact is that it is an intrinsic limit, which does not require a choice of coordinate systems on X_i. By the work of Cheeger, Colding, and Tian [2002], we know X_∞ can be written as the union of the regular part $\mathcal{R}$, which is a smooth open manifold endowed with a $C^{1,\alpha}$ Kähler structure $(J_\infty, \omega_\infty)$, and the singular set $\mathcal{S}$, which has Hausdorff codimension at least 4. In the case when $n = 2$ it has been known by Anderson [1989], Bando, Kasue, and Nakajima [1989], and Tian [1990b] that X_∞ has only isolated orbifold singularities. By possibly passing to a further subsequence, we may assume the Chern connection on L_i converges modulo gauge transformations to the Chern connection of a hermitian holomorphic line bundle $L_{\mathcal{R}}$ on $\mathcal{R}$. The following result establishes a basic connection with algebraic geometry

Theorem 2.1 (Donaldson and Sun [2014]). *(1) X_∞ is naturally homeomorphic to a normal projective variety in such a way that the algebraic singularities are a subset of the metric singularities, and, by passing to subsequence the convergence of X_i to X_∞ can be realized in a fixed Hilbert scheme.*

(2) If furthermore we assume ω_i is Kähler–Einstein, then the two singular sets are equal, and the algebraic singularities of X_∞ are log terminal in the sense of minimal model program. Moreover, ω_∞ extends to a global Kähler current that satisfies a singular Kähler–Einstein equation in the sense of pluripotential theory, c.f. Eyssidieux, Guedj, and Zeriahi [2009].

Remark 2.2.

- The projective algebraic structure on X_∞ is intrinsically determined by $(\mathcal{R}, L_{\mathcal{R}})$. Namely, denote by $\iota : \mathcal{R} \to X_\infty$ the natural inclusion map, and we can define $\mathcal{O}_{X_\infty} := \iota_*\mathcal{O}_{\mathcal{R}}$ and $L_\infty := \iota_* L_{\mathcal{R}}$. Then the precise meaning of (1) is that $(X_\infty, \mathcal{O}_{X_\infty})$ is a normal complex analytic space and L_∞ is an ample $\mathbb{Q}$-line bundle on X_∞. Moreover, by passing to a subsequence, (X_i, L_i) and (X_∞, L_∞) can be fit into a *flat* family.
- In the Kähler–Einstein case with $\lambda = 1$, the line bundle L_i is given by $K_{X_i}^{-1}$, and the diameter bound is automatically satisfied by Bonnet–Myers's theorem. In this case X_∞ is a smoothable $\mathbb{Q}$-*Fano variety*. Theorem 2.1 thus provides a *topological*

compactification of the moduli space of Kähler–Einstein Fano manifolds in each dimension, by adding all the possible Gromov–Hausdorff limits.

A key ingredient in the proof of Theorem 2.1 is to obtain the *partial C^0 estimate*, conjectured by G. Tian in ICM 1990 (c.f. Tian [1990a]). Recall, in general that if (X, L) is a polarized Kähler manifold and ω is a Kähler metric in $2\pi c_1(L)$, then ω determines a hermitian metric $|\cdot|$ on L, unique up to constant multiple. For all k we have an induced L^2 norm $\|\cdot\|$ on $H^0(X, L^k)$ (defined with respect to the volume form of $k\omega$). To compare these we define the *density of state function* (or the *Bergman function*)

$$\rho_{X,\omega,k}(x) = \sup_{s\in H^0(X,L^k),s\neq 0} \frac{|s(x)|^2}{\|s\|^2}.$$

By the Kodaira embedding theorem this is a positive function for k sufficiently large. It is easy to see for each $x \in X$, the supremum is achieved on a unique one dimension subspace $\mathbb{C}_x \subset H^0(X, L^k)$, and $\rho_{X,\omega,k}$ is a smooth function on X. Moreover, we have a C^∞ asymptotic expansion (c.f. Zelditch [1998])

$$\rho_{X,\omega,k} = 1 + \frac{S(\omega)}{2}k^{-1} + \cdots . \tag{2-1}$$

Its importance can be seen as follows

- Denote $N_k = \dim H^0(X, L^k)$, and choose an L^2 orthonormal basis $\{s_1, \cdots, s_{N_k}\}$ of $H^0(X, L^k)$, then there is an alternative expression

$$\rho_{X,\omega,k}(x) = \sum_{i=1}^{N_k} |s_i(x)|^2.$$

  So we get $N_k = \int_X \rho_{X,\omega,k}(k\omega)^n/n!$, and (2-1) can be viewed as a local version of the Riemann–Roch formula.

- Using an orthonormal basis of $H^0(X, L^k)$, for k large we get an embedding $\iota_k : X \to \mathbb{P}^{N_k-1}$, unique up to unitary transformations. Then we have

$$\iota_k^*\omega_k = k\omega + i\partial\bar\partial \log \rho_k. \tag{2-2}$$

  So by (2-1) we know $k^{-1}\iota_k^*\omega_k$ converges smoothly to ω as k tends to infinity. This is the *Kähler quantization* picture, which is essential in studying the relationship between constant scalar curvature Kähler metrics and algebraic stability (c.f. Donaldson [2001]).

Theorem 2.3 (Donaldson and Sun [2014]). *Under the hypothesis in the beginning of this section, there are k_0 and $\epsilon_0 > 0$ that depend only on n, Λ and D, such that $\rho_{X_i,\omega_i,k_0} \geq \epsilon_0$ for all i.*

Remark 2.4. (1). When $n = 2$ this is proved by Tian [1990b], using the fact that in this case the Gromov–Hausdorff limit is an orbifold; for general n this is conjectured by Tian [1990a]. Indeed Tian's original conjecture is stated under more general assumptions, and there has been substantial progress on this, see for example Theorem 1.9 in Chen and B. Wang [2014].

(2). One may ask if there is a uniform asymptotic behavior of the density of state function as k tends to infinity. In general the expansion (2-1) can not hold uniformly independent of i, but a weaker statement is plausible, see Conjecture 5.15 in Donaldson and Sun [2014].

(3). It follows from Theorem 2.3 and (2-2) that the embedding map ι_{k_0} has a uniform Lipschitz bound for all i, hence one can pass to limit and obtain a Lipschitz map from the Gromov–Hausdorff limit X_∞ to a fixed projective space. This is the starting point to prove Theorem 2.1.

The proof of Theorem 2.3 is based on Hörmander's construction of holomorphic sections on definite powers of L_i with control. The idea is to first find holomorphic sections of Gaussian type in a local model, then graft them to the manifolds X_i to get approximately holomorphic sections, and finally correct these to genuine holomorphic sections by solving a $\bar\partial$ equation. The solvability of $\bar\partial$ equation with uniform estimates only uses one global geometric assumption, namely the lower bound of Ricci curvature of ω_i.

Here we briefly describe the notion of model Gaussian sections. Suppose first we are at a *smooth* point $x \in X_i$ for some *fixed* i. If we dilate the metric ω_i to $k\omega_i$ based at x, then as k tends to infinity we get in the limit the standard flat metric on $\mathbb{C}^n$. The dilation has the effect of replacing the line bundle L_i by L_i^k and it is also important to notice that the corresponding limit line bundle is the trivial holomorphic bundle on $\mathbb{C}^n$ endowed with the non-trivial hermitian metric $e^{-|z|^2/2}$ whose curvature is exactly the flat metric. The obvious trivial section is then naturally a Gaussian section. Using this one can construct for k large a holomorphic section of L_i^k over X_i, whose L^2 norm has a definite upper bound and its pointwise norm at x has a definite positive lower bound. This implies $\rho_{X_i,\omega_i,k}(x) \geq \epsilon > 0$. The difficulty in the proof of Theorem 2.3 is then to obtain uniform estimate on both k and ϵ, which is not a priori clear since it is conceivable that when the metric ω_i starts to form singularities the region where the Gaussian model behaves well for a fixed k shrinks to one point. For this purpose we use the Gromov–Hausdorff limit X_∞ and consider a sequence of points $x_i \in X_i$ that tend to a point $x_\infty \in X_\infty$. If x_∞ is a smooth point then previous argument goes through with little change to yield $\rho_{X_i,\omega_i,k}(x_i) \geq \epsilon > 0$ for k and ϵ independent of i. If x_∞ is singular, then we can dilate the metric on X_∞ based at x_∞. By passing to a subsequence we can obtain a *pointed* Gromov–Hausdorff limit, called a *tangent cone*. It is a *metric cone* $C(Y)$ over a compact length space Y, which in particular admits a dilation action.

By Cheeger, Colding, and Tian [2002] we also know that $C(Y)$ is a $C^{1,\alpha}$ Kähler manifold outside a closed subset of Hausdorff codimension at least 4 which is invariant under dilation. On this smooth part, being a metric cone implies that the metric can be written as $g = \frac{1}{2}\mathrm{Hess}(r^2)$, where r is the distance to the cone vertex. Now the crucial point in our Kähler setting is that the Kähler form can correspondingly be written as $\omega = \frac{1}{2} i\partial\bar{\partial} r^2$, so in particular it is the curvature of the trivial holomorphic bundle with hermitian metric $e^{-r^2/2}$. This is exactly the generalization of the above local model on $\mathbb{C}^n$. Thus we also have the candidate Gaussian section in this case, but there are various technical difficulties that had to be overcome in Donaldson and Sun [2014], mainly due to the appearance of singularities on the tangent cones $C(Y)$. One point to notice is that the proof does *not* require the tangent cone at x_∞ to be unique, even though this turns out to be true and we will discuss more in the next section.

We mention that the extension of Theorem 2.1 and 2.3 to the case of Kähler–Einstein metrics with cone singularities plays a key role in the proof of the Yau–Tian–Donaldson conjecture for Fano manifolds (c.f. Chen, Donaldson, and Sun [2015a,b,c]). There are also other applications to the study of Gromov–Hausdorff limits of Kähler–Einstein metrics in the Calabi–Yau and general type case, c.f. Rong and Zhang [2011], Tosatti [2015], and Song [2017]. In both cases the diameter bound (hence the volume non-collapsing condition) is not automatically satisfied, and is essentially equivalent to the algebraic limit having at worst log terminal singularities.

There are a few interesting directions that require further development:

(1). Prove a local version of Theorem 2.1. Namely, let X_∞ be a Gromov–Hausdorff limit of a sequence of (incomplete) Kähler manifolds (X_i, ω_i) with diameter 1, uniformly bounded Ricci curvature, and satisfying a uniform volume non-collapsing condition, can we prove X_∞ is naturally a complex analytic space?

Notice in general one can not expect to fit X_i and X_∞ into a flat family in the absence of the line bundles, since one can imagine certain holomorphic cycles being contracted under the limit process. The answer to the above question will have applications in understanding convergence of Calabi–Yau metrics when Kähler class becomes degenerate, see Collins and Tosatti [2015]. In a related but different context, G. Liu [2016] studied the case when the Ricci curvature bound is replaced by a lower bound on the bisectional curvature, and used it to make substantial progress towards Yau's uniformization conjecture for complete

Kähler manifolds with non-negative bisectional curvature. Most generally, one would expect that in the above question to draw the complex-analytic consequence the most crucial assumption is a uniform lower bound on Ricci curvature, and the upper bound on Ricci curvature is more related to the regularity of the limit metrics.

(2). In general collapsing is un-avoidable. Can we describe the complex/ algebro-geometric meaning of the Gromov–Hausdorff limits?

Collapsing with uniformly bounded Riemannian sectional curvature has been studied extensively in the context of Riemannian geometry through the work of Cheeger–Fukaya–Gromov and others. In general one expects a structure of fibrations (in the generalized sense). Even in this case the above question is not well-understood. In general with only Ricci curvature bound there has been very few results on the regularity of the limit space itself (see Cheeger and Tian [2006] for results when $n = 2$).

Ultimately one would like to understand the case of constant scalar curvature or extremal Kähler metrics, where we are currently lacking the analogous foundations of the Cheeger–Colding theory which depends on comparison geometry of Ricci curvature, and so far we only have results in the non-collapsed case, see for example Tian and Viaclovsky [2008], Chen and Weber [2011].

3 Singularities

In this section we focus on finer structure of the Gromov–Hausdorff convergence studied in Theorem 2.1, and restrict to the case when (X_i, ω_i) is Kähler–Einstein. It is a folklore picture that when singularities occur certain non-compact Ricci-flat spaces must bubble off. To be more precise, suppose $p \in X_\infty$ is a singular point, and $p_i \in X_i$ is a sequence of points that converge to p. Take any sequence of integers $k_i \to \infty$ and consider the rescaled spaces $(X_i, L_i^{k_i}, k_i\omega_i, p_i)$, then by passing to subsequence, we always get a *pointed Gromov–Hausdorff limit* (Z, p_∞), which is a non-compact metric space.

Theorem 3.1 (Donaldson and Sun [2017])**.** *Any such limit Z is naturally a normal affine algebraic variety which admits a singular Ricci-flat Kähler metric.*

To explain the meaning of this, similar to Remark 2.2, we know the complex-analytic structure on Z is determined by the regular part of Z (in the sense of Cheeger, Colding, and Tian [2002]). To understand intrinsically the affine structure, we denote by $R(Z)$ the ring of holomorphic functions on Z that grow at most polynomially fast at infinity, then it is proved in Donaldson and Sun [2017] that $R(Z)$ is finitely generated and $\mathrm{Spec}(R(Z))$ is complex-analytically isomorphic to Z.

In dimension two, Z is an ALE Ricci-flat space possibly with orbifold singularities. In higher dimensions, by the volume non-collapsing condition we know Z is *asymptotically* *conical*. There has been extensive study on these spaces in the case when the tangent cone at infinity is smooth, see for example van Coevering [2011], Conlon and Hein [2015], C. Li [2015].

Let $\mathfrak{B}$ be the set of all *bubbles* at p, i.e. the set of pointed Gromov–Hausdorff limits of $(X_i, k_i\omega_i, p_i)$ with $p_i \to p$ and $k_i \to \infty$. It is an interesting question to find complex/algebro-geometric characterization of this set, which potentially forms a *bubble tree* structure at p. For each $Z \in \mathfrak{B}$, the Bishop–Gromov volume comparison defines an invariant $v(Z)$, namely, the asymptotic volume ratio of Z

$$v(Z) = \lim_{r \to \infty} Vol(B(p_\infty, r))r^{-2n}$$

A special element in $\mathfrak{B}$ is a metric tangent cone $C(Y)$ at p, which has the *smallest* asymptotic volume ratio among all the bubbles at p. The metric cone structure imposes an extra dilation symmetry on $C(Y)$, and this has the corresponding algebro-geometric meaning

Theorem 3.2 (Donaldson and Sun [2017]). *A tangent cone $C(Y)$ is naturally a normal affine algebraic cone.*

This requires some explanation. Let $R(C(Y))$ be the affine coordinate ring of $C(Y)$. On the regular part of $C(Y)$, we have a *Reeb* vector field $\xi = Jr\partial_r$ which is a holomorphic Killing vector field. It generates holomorphic isometric action of a compact torus $\mathbb{T}$ on $C(Y)$. This action induces a weight space decomposition of $R(C(Y))$, which can be written as

$$R(C(Y)) = \bigoplus_{\mu \in \mathbb{H}} R_\mu(C(Y)).$$

Here $R_\mu(C(Y))$ is the space of homogeneous holomorphic functions f on $C(Y)$ satisfying $\mathcal{L}_\xi f = i\mu f$ (i.e. homogeneous of degree μ), and the *holomorphic spectrum* $\mathbb{H}$ is the set of all $\mu \in \mathbb{R}_{\geq 0}$ such that $R_\mu(C(Y)) \neq 0$. In general we know $\mathbb{H}$ is contained in the set of algebraic numbers, but not necessarily a subset of $\mathbb{Q}$. This *positive* $\mathbb{R}$-grading on $R(C(Y))$ is the precise meaning for $C(Y)$ to be an affine algebraic *cone* in Theorem 3.2.

Theorem 3.3 (Donaldson and Sun [ibid.]). *Given any $p \in X_\infty$, there is a unique metric tangent cone at p, as an affine algebraic cone endowed with a singular Ricci-flat Kähler metric.*

One step in the proof is to show that the holomorphic spectrum $\mathbb{H}$ is a priori unique. The crucial observation is that the set of all possible tangent cones at p form a connected and compact set under the pointed Gromov–Hausdorff topology, and $\mathbb{H}$ consists of only algebraic numbers so must be rigid under continuous deformations. The latter uses the volume minimization principle of Martelli, Sparks, and Yau [2008] in Sasaki geometry. Recall that $C(Y)$ being Ricci-flat is equivalent to Y being *Sasaki–Einstein* possibly with singularities. Consider the open convex cone $\mathcal{C}$ in $Lie(\mathbb{T})$ consisting of elements η which also induce positive gradings on $R(C(Y))$. Then by Martelli, Sparks, and Yau [ibid.] we know ξ is the unique critical point of the (suitably normalized) *volume function* on $\mathcal{C}$, which is a convex rational function with rational coefficients (noticing there is a natural rational structure in $Lie(\mathbb{T})$). This fact leads to the algebraicity of $\mathbb{H}$.

To describe the relationship between the local algebraic singularity of X_∞ at p and the tangent cone $C(Y)$, we define a *degree* function on $\mathcal{O}_p$, the local ring of germs of holomorphic functions at p, by setting

$$d(f) = \lim_{r\to 0} \frac{\sup_{B(p,r)} \log|f|}{\log r}$$

By Donaldson and Sun [2014], for all nonzero f, $d(f)$ is always well-defined and it belongs to $\mathbb{H}$. Indeed d defines a *valuation* on $\mathcal{O}_p$. So we can define a graded ring associated to this

$$R_p = \bigoplus_{\mu\in\mathbb{H}} \{f \in \mathcal{O}_p | d(f) \geq \mu\}/\{f \in \mathcal{O}_p | d(f) > \mu\}.$$

Theorem 3.4 (Donaldson and Sun [2017]).

- *R_p is finitely generated and $Spec(R_p)$ defines a normal affine algebraic cone W, which can be realized as a weighted tangent cone of (X_∞, p) under a local complex analytic embedding into some affine space.*
- *There is an equivariant degeneration from W to $C(Y)$ as affine algebraic cones.*

Remark 3.5. The finite generation of R_p and the fact that W is normal put on very strong constraint on the valuation d. The proof depends on Theorem 3.2 and a three-circle type argument that relates elements in $\mathcal{O}_p$ and $R(C(Y))$. In the context of Theorem 3.1 there is a similar result relating Z and its tangent cone at *infinity*.

This situation is analogous to the Harder–Narasimhan–Seshadri filtration for unstable holomorphic vector bundles. It also suggests a *local* notion of stability for algebraic singularities, since using the extension of the Yau–Tian–Donaldson conjecture to affine algebraic cones by Collins and Székelyhidi [2012], one can view $C(Y)$ as a K-stable algebraic cone and W as a K-semistable algebraic cone.

One interesting point is that W is only an algebraic variety but does not support a natural metric structure. In Donaldson and Sun [2017] we conjectured that W and $C(Y)$ are both invariants of $\mathcal{O}_p$ and there should a purely algebro-geometric way of characterizing

W. C. Li [2015] made this conjecture more precise by giving an algebro-geometric interpretation of the volume of an affine cone, and formulating a corresponding conjecture that the valuation d should be the unique one that minimizes volume. This brings interesting connections with earlier work on asymptotic invariants and K-stability in algebraic geometry. It is an extension of the Martelli–Sparks–Yau volume minimization principle. There is much progress in this direction, see Blum [2016], C. Li and Y. Liu [2016], and C. Li and Xu [2017].

One of the motivation for studying metric tangent cones is related to precise understanding of the metric behavior of a singular Kähler–Einstein metric. The following result, which uses the result of Fujita [2015] and C. Li and Y. Liu [2016], gives the first examples of compact Ricci-flat spaces with isolated conical singularities.

Theorem 3.6 (Hein and Sun [2017])**.** *Let (X, L) be a $\mathbb{Q}$-Gorenstein smoothable projective Calabi–Yau variety with isolated canonical singularities, each locally complex-analytically isomorphic to a strongly regular affine algebraic cone which admits a Ricci-flat Kähler cone metric, then there is a unique singular Calabi–Yau metric $\omega \in 2\pi c_1(L)$ which is smooth away from the singular locus of X, and at each singularity is asymptotic to the Ricci-flat Kähler cone metric at a polynomial rate.*

Remark 3.7.

- The notion of *strong regularity* is a technical assumption which is equivalent to that the affine algebraic cone coincides with its Zariski tangent cone at the vertex.
- An important special case is when the singularities of X are ordinary double points, in which case the Ricci-flat Kähler cone metric can be explicitly written down, and Theorem 3.6 also implies the existence of *special lagrangian vanishing spheres* on a generic smoothing of X.

The above general strategy has other applications, one is to the study of the asymptotic behavior of geometric flows, as we shall describe in Section 5, and the other is to the study of singularities of Hermitian-Yang–Mills connections (c.f. Chen and Sun [2017]).

4 Moduli spaces

Theorem 2.1 gives a *Gromov–Hausdorff compactification* of the moduli space of Kähler–Einstein Fano manifolds in each fixed dimension, as a topological space. By Theorem 1.1 this is the same as a compactification of the moduli space of K-stable Fano manifolds, so it is natural to ask for algebro-geometric meaning of this moduli space itself; furthermore one would like to characterize them explicitly since there are many concrete examples of families of smooth Fano manifolds. Understanding this would also lead to new examples of K-stable Fano varieties, including singular ones.

In dimension two, there are only four families of Fano manifolds with non-trivial moduli. They are *del Pezzo surfaces* of anti-canonical degree $d \in \{1, 2, 3, 4\}$, i.e. the blown-ups of $\mathbb{CP}^2$ at $9 - d$ points in very general position. Let $\mathfrak{M}_d$ be the Gromov–Hausdorff compactification of moduli space of Kähler–Einstein metrics on del Pezzo surfaces of degree d.

Theorem 4.1 (Odaka, Spotti, and Sun [2016]). *Each $\mathfrak{M}_d$ is naturally homeomorphic to an explicitly constructed moduli space $\mathfrak{M}_d^{alg}$ of certain del Pezzo surfaces with orbifold singularities.*

The construction of $\mathfrak{M}_d^{alg}$ depends on the classical geometry of del Pezzo surfaces, whose moduli is closely related to geometric invariant theory.

- Objects in $\mathfrak{M}_3^{alg}$ are GIT stable[1] cubics in $\mathbb{P}^3$. These were classified by Hilbert.
- Objects in $\mathfrak{M}_4^{alg}$ are GIT stable complete intersections of two quadrics in $\mathbb{P}^4$. These were classified by Mabuchi and Mukai [1993], who also proved Theorem 4.1 in this case, with a more involved argument.
- Objects in $\mathfrak{M}_2^{alg}$ are either double covers of $\mathbb{P}^2$ branched along a GIT stable quartic curve, or double covers of the weighted projective plane $\mathbb{P}(1, 1, 4)$ branched along a curve of the form $z^2 = f_8(x, y)$ for a GIT stable octic in 2 variables. This moduli was constructed by Mukai [1995].
- Objects in $\mathfrak{M}_1^{alg}$ are more complicated to describe, but a generic element is a double cover of $\mathbb{P}(1, 1, 2)$ branched along a sextic curve that is GIT stable in a suitable sense (even though the automorphism group of $\mathbb{P}(1, 1, 2)$ is not reductive). $\mathfrak{M}_1^{alg}$ is a certain birational modification of this GIT moduli space.

We briefly describe the general strategy in the proof, which also applies in higher dimensions (see Theorem 4.2). For a more detailed survey see Spotti [n.d.].

- Show $\mathfrak{M}_d$ is non-empty. This can be achieved by studying a special element in each family, for example through computation of α-invariant (see Tian and Yau [1987]), or alternatively Arezzo, Ghigi, and Pirola [2006], or by a glueing construction.
- Rough classification of Gromov–Hasudorff limits. The Bishop–Gromov volume comparison yields that at any singularity of the form $\mathbb{C}^2/\Gamma$, we have $|\Gamma| < 12/d$. For larger d we get stronger control on Γ hence the corresponding $\mathfrak{M}_d$ is simpler. Using this one can estimate the *Gorenstein index* of the Gromov–Hausdorff limits,

[1] The stability in this article means *polystablity* in the usual literature.

and then understand their anti-canonical geometry. This is related to the explicit determination of the smallest number k_0 in Theorem 2.3.

By R. J. Berman [2016] all the objects X in $\mathfrak{M}_d$ are K-stable, and a crucial ingredient in Odaka, Spotti, and Sun [2016] is that one can often compare K-stability with GIT stability. This implies X is also GIT stable in appropriate sense, and suggests that $\mathfrak{M}_d$ is closely related to GIT moduli.

- Construction of $\mathfrak{M}_d^{alg}$ so that the natural map from $\mathfrak{M}_d$ to $\mathfrak{M}_d^{alg}$ is a homeomorphism. The key point is to make sure $\mathfrak{M}_d^{alg}$ is Hausdorff and the map is well-defined, i.e. any possible Gromov–Hausdorff limit in $\mathfrak{M}_d$ is included in $\mathfrak{M}_d^{alg}$. In general one can start with a natural GIT moduli space, and perform birational modifications that are suggested by the previous step. For example when $d = 2$, the natural GIT moduli of quartic curves contains a point that corresponds to a reducible surface, which we know can not be in $\mathfrak{M}_d$; this motivates one to blow-up the point and the exceptional divisor turns out to correspond to a different GIT as described above.

Theorem 4.2.

- *Spotti and Sun [2017]: For all n, the Gromov–Hausdorff compactification of the moduli space of Kähler–Einstein metrics on complete intersection of two quadrics in $\mathbb{P}^{n+2}$ is naturally homeomorphic to the GIT moduli space.*
- *Y. Liu and Xu [2017]: The Gromov–Hausdorff compactification of the moduli space of Kähler–Einstein metrics on cubics in $\mathbb{P}^4$ is naturally homeomorphic to the GIT moduli space.*

A major step is to prove the Gromov–Hausdorff limit has a well-defined canonical line bundle, i.e., is *Gorenstein*, and has large divisibility. This relies on the following deep theorem of K. Fujita proved via K-stability, and the later generalization by Y. Liu [2016] to give an improvement of the Bishop–Gromov volume comparison.

Theorem 4.3.

- *Fujita [2015]: Let X be a Kähler–Einstein Fano manifold in dimension n, then $(-K_X)^n \leq (n+1)^n$, and equality holds if and only if X is isomorphic to $\mathbb{P}^n$.*

It is very likely that similar results to Theorem 4.2 can be established for most families of Fano threefolds, and some classes of higher dimensional Fano manifolds with large anti-canonical volume. There is a related conjecture on a local analogue of Theorem 4.3. For more on this see Spotti and Sun [2017].

Now we move on to discuss general abstract results in higher dimension concerning moduli space of Kähler–Einstein/K-stable manifolds. As applications of Chen, Donaldson, and Sun [2015a,b,c] we have

- Odaka [2012a] and Donaldson [2015]: The moduli space of K-stable Fano manifolds with discrete automorphism group is Zariski open.
- Spotti, Sun, and Yao [2016]: A smoothable $\mathbb{Q}$-Fano variety admits a singular Kähler–Einstein metric if and only if it is K-stable.
- C. Li, X. Wang, and Xu [2014] and Odaka [2015]: The Gromov–Hausdorff compactification of moduli space of Kähler–Einstein Fano manifolds in a fixed dimension is naturally a proper separated algebraic space.

We remark that one important technical aspect is still open, namely, the *projectivity* of the moduli space. There is a well-defined CM line bundle on the moduli space. Over the locus parametrizing smooth Fano manifolds it admits a natural Wei–Peterson metric of positive curvature, and this locus has been shown to be quasi-projective C. Li, X. Wang, and Xu [2015].

There is also recent progress in the general type case. Here we already have a compactification using minimal model program, namely, the *KSBA moduli space*, where the boundary consists of varieties with semi-log-canonical singularities (in one dimension it is the same as being nodal). Odaka [2013, 2012b] proved that these are exactly the ones which are K-stable, and R. J. Berman and Guenancia [2014] established the existence of a unique singular Kähler–Einstein metric in a suitable weak sense. It is further shown by Song [2017] that under a KSBA degeneration, the Kähler–Einstein metric on smooth fibers converges in the pointed Gromov–Hausdorff sense to the metric completion of the log terminal locus on the central fiber. So far the proof uses deep results in algebraic geometry but one certainly hopes for a more differential-geometric theory in the future.

5 Optimal degenerations

Let (X, L, ω) be a polarized Kähler manifold with $[\omega] = 2\pi c_1(L)$. In this section we focus on two natural geometric flows emanating from ω. We first consider the *Ricci flow*

$$\frac{\partial}{\partial t}\omega(t) = \omega(t) - Ric(\omega(t)) \tag{5-1}$$

and restrict to the case when X is Fano and $L = K_X^{-1}$. This case is immediately related to the Yau–Tian–Donaldson conjecture; the general case is also interesting and is related to the analytic minimal model program, but is beyond the scope of this article.

Clearly a fixed point of (5-1) is exactly a Kähler–Einstein metric. There are also self-similar solutions to (5-1), i.e., solutions $\omega(t)$ that evolve by holomorphic transformations of X. These correspond to *Ricci solitons*, which are governed by the equation $Ric(\omega) = \omega + \mathfrak{L}_V\omega$, where V is a holomorphic vector field on X.

It is well-known that in our case a smooth solution $\omega(t)$ exists for all time $t > 0$ with $[\omega(t)] = 2\pi c_1(K_X^{-1})$. A folklore conjecture, usually referred to as the *Hamilton–Tian conjecture*, states that as $t \to \infty$ by passing to subsequence one should obtain Gromov–Hausdorff limits which are Ricci solitons off a singular set of small size. This is now confirmed by Chen and B. Wang [2014].

Theorem 5.1 (Chen and B. Wang [ibid.]). *As $t \to \infty$, by passing to a subsequence $(X, \omega(t))$ converges in the Gromov–Hausdorff sense to a $\mathbb{Q}$-Fano variety endowed with a singular Kähler–Ricci soliton metric $(X_\infty, V_\infty, \omega_\infty)$.*

This can be viewed as a generalization of the Cheeger–Colding theory and the results in Section 2 to the parabolic case. The proof makes use of the deep results of Perelman. In connecting with algebraic geometry we have

Theorem 5.2 (Chen, Sun, and B. Wang [2015]).

- *As $t \to \infty$, there is a unique Gromov–Hausdorff limit $(X_\infty, V_\infty, \omega_\infty)$.*
- *If X is K-stable, then X_∞ is isomorphic to X, $V_\infty = 0$, and ω_∞ is a smooth Kähler–Einstein metric on X_∞.*
- *If X is K-unstable, then the flow $\omega(t)$ defines a unique degeneration of X to a $\mathbb{Q}$-Fano variety $\bar{X}$ with a holomorphic vector field $\bar{V}$, and there is an equivariant degeneration from $(\bar{X}, \bar{V})$ to (X_∞, V_∞).*

The second item gives an alternative proof of the Yau–Tian–Donaldson conjecture for Fano manifolds. The third item requires slightly more explanation. The flow $\omega(t)$ induces a family of L^2 norms $\|\cdot\|_t$ on $H^0(X, K_X^{-k})$ for all k, and yields a notion of *degree* of a section $s \in H^0(X, L^k)$ by setting

$$d(s) = \lim_{t\to\infty} \frac{1}{t} \log \|s\|_t.$$

The main point in Chen, Sun, and B. Wang [ibid.] is that this is well-defined and gives a filtration of the homogeneous coordinate ring of (X, L); moreover, the associated graded ring defines a normal projective variety $\bar{X}$, and the grading determines a holomorphic vector field $\bar{V}$. So $(\bar{X}, \bar{V})$ is canonically determined by the flow $\omega(t)$, hence by the initial metric ω.

It is conjectured in Chen, Sun, and B. Wang [ibid.] that when X is K-unstable, both $(\bar{X}, \bar{V})$ and $(X_\infty, V_\infty, \omega_\infty)$ are invariants of X itself, i.e. are independent of the initial metric ω, and $(\bar{X}, \bar{V})$ defines a unique *optimal degeneration* of X. By W. He [2016], R. J. Berman and Nystrom [2014], and Dervan and Székelyhidi [2016] we know a partial answer to this, and it seems possible to eventually obtain a purely algebro-geometric characterization.

Now we turn to discuss the *Calabi flow*, starting from an initial metric ω with $[\omega] = 2\pi c_1(L)$. The equation takes the form

$$\frac{\partial}{\partial t}\omega(t) = -\Delta_{\bar{\partial}} Ric(\omega(t))(= -i\bar{\partial}\partial S(\omega(t))) \tag{5-2}$$

A fixed point is a Kähler metric with constant scalar curvature, and a self-similar solution is an extremal Kähler metric. From the viewpoint of the infinite dimensional moment map picture of Fujiki–Donaldson, the Calabi flow has an interesting geometric meaning, at least on a formal level. First it is the negative gradient flow of a geodesically convex functional (the *Mabuchi functional*) on the Kähler class $\mathcal{H}$, so it decreases a natural distance function on $\mathcal{H}$ (c.f. Calabi and Chen [2002]). If we transform it to a flow of integrable almost complex structures compatible with a fixed symplectic form then it is also the negative gradient flow of the *Calabi functional*.

A main difficulty in the study of Calabi flow (c.f. Chen and W. Y. He [2008]) arises from the fact it is a *fourth* order geometric evolution equation and the usual technique of maximum principle does not apply directly. There has been little progress on the problem of general long time existence. If we put aside all the analytic difficulties, then we do have a picture on the asymptotic behavior of the flow, analogous to Theorem 5.2.

Theorem 5.3 (Chen, Sun, and B. Wang [2015]). *Let $\omega(t)(t \in [0, \infty))$ be a smooth solution of the Calabi flow in the class $2\pi c_1(L)$, and assume the Riemannian curvature of $\omega(t)$ and the diameter are uniformly bounded for all t. Then*

- *There is a unique Gromov–Hausdorff limit $(X_\infty, L_\infty, V_\infty, \omega_\infty)$, where ω_∞ is a smooth extremal Kähler metric on a smooth projective variety X_∞ with $[\omega_\infty] = 2\pi c_1(L_\infty)$, and $V_\infty = \nabla S(\omega_\infty)$ is a holomorphic vector field.*
- *If (X, L) is K-stable, then (X_∞, L_∞) is isomorphic to (X, L), $V_\infty = 0$, and ω_∞ is a constant scalar curvature Kähler metric on X.*
- *If (X, L) is K-unstable, then it gives rise to an optimal degeneration of (X, L) to $(\bar{X}, \bar{L}, \bar{V})$, which minimizes the normalized Donaldson–Futaki invariant, and there is an equivariant degeneration from $(\bar{X}, \bar{L}, \bar{V})$ to $(X_\infty, L_\infty, V_\infty)$.*

There is also a generalized statement for extremal Kähler metrics Chen, Sun, and B. Wang [ibid.]. Notice in general we should not expect the Calabi flow to satisfy the strong geometric hypothesis in Theorem 5.3. First, the curvature may blow up and singularities can form, similar to the case of Ricci flow; second, the diameter can go to infinity (c.f. Székelyhidi [2009]) and collapsing may happen. The second issue is related to the folklore expectation that one may need to strengthen the notion of K-stability to certain *uniform*

K-stability in the statement of Yau–Tian–Donaldson conjecture. Nevertheless we expect Theorem 5.3 will lead to existence results of extremal Kähler metrics in concrete cases.

There are also recent progress on analytic study of the asymptotic behavior on the Calabi flow, related to the conjectural picture described by Donaldson [2004]. On one hand, in complex dimension two, H. Li, B. Wang, and Zheng [2015] proved that, assuming the existence of a constant scalar curvature Kähler metric ω_0 in the class $2\pi c_1(L)$, if a Calabi flow $\omega(t)$ in $2\pi c_1(L)$ exists for $t \in [0, \infty)$, then as $t \to \infty$ the flow must converge to ω_0, modulo holomorphic transformations of X. On the other hand, R. J. Berman, Darvas, and Lu [2017] proved a dichotomy for the behavior of *weak solution* to the Calabi flow in the sense of Streets [2014], to the effect that it either diverges to infinity with respect to a natural distance on $\mathcal{H}$, or it converges to a weak minimizer of the Mabuchi functional in a suitable sense. Finally, in Chen and Sun [2014] Calabi flow on a small complex structure deformation of a constant scalar curvature Kähler manifold is studied, which also leads to a generalized uniqueness result.

References

Michael T. Anderson (1989). "Ricci curvature bounds and Einstein metrics on compact manifolds". *J. Amer. Math. Soc.* 2.3, pp. 455–490. MR: 999661 (cit. on p. 1014).

Claudio Arezzo, Alessandro Ghigi, and Gian Pietro Pirola (2006). "Symmetries, quotients and Kähler-Einstein metrics". *J. Reine Angew. Math.* 591, pp. 177–200. MR: 2212883 (cit. on p. 1022).

Thierry Aubin (1976). "Équations du type Monge-Ampère sur les variétés kähleriennes compactes". *C. R. Acad. Sci. Paris Sér. A-B* 283.3, Aiii, A119–A121. MR: 0433520 (cit. on p. 1012).

Shigetoshi Bando, Atsushi Kasue, and Hiraku Nakajima (1989). "On a construction of coordinates at infinity on manifolds with fast curvature decay and maximal volume growth". *Invent. Math.* 97.2, pp. 313–349. MR: 1001844 (cit. on p. 1014).

Robert J. Berman (2016). "K-polystability of $\mathbb{Q}$-Fano varieties admitting Kähler-Einstein metrics". *Invent. Math.* 203.3, pp. 973–1025. MR: 3461370 (cit. on pp. 1012, 1023).

Robert J. Berman and Bo Berndtsson (2017). "Convexity of the K-energy on the space of Kähler metrics and uniqueness of extremal metrics". *J. Amer. Math. Soc.* 30.4, pp. 1165–1196. MR: 3671939 (cit. on p. 1013).

Robert J. Berman, Tamás Darvas, and Chinh H. Lu (2017). "Convexity of the extended K-energy and the large time behavior of the weak Calabi flow". *Geom. Topol.* 21.5, pp. 2945–2988. MR: 3687111 (cit. on p. 1027).

Robert J. Berman and Henri Guenancia (2014). "Kähler-Einstein metrics on stable varieties and log canonical pairs". *Geom. Funct. Anal.* 24.6, pp. 1683–1730. MR: 3283927 (cit. on p. 1024).

Robert J. Berman and David Witt Nystrom (Jan. 2014). "Complex optimal transport and the pluripotential theory of Kähler-Ricci solitons". arXiv: 1401.8264 (cit. on p. 1025).

Robert Berman, Sébastien Boucksom, and Mattias Jonsson (Sept. 2015). "A variational approach to the Yau-Tian-Donaldson conjecture". arXiv: 1509.04561 (cit. on p. 1012).

Harold Blum (June 2016). "Existence of valuations with smallest normalized volume". arXiv: 1606.08894 (cit. on p. 1021).

E. Calabi and Xiuxiong Chen (2002). "The space of Kähler metrics. II". *J. Differential Geom.* 61.2, pp. 173–193. MR: 1969662 (cit. on p. 1026).

J. Cheeger, T. H. Colding, and G. Tian (2002). "On the singularities of spaces with bounded Ricci curvature". *Geom. Funct. Anal.* 12.5, pp. 873–914. MR: 1937830 (cit. on pp. 1014, 1017, 1018).

Jeff Cheeger and Gang Tian (2006). "Curvature and injectivity radius estimates for Einstein 4-manifolds". *J. Amer. Math. Soc.* 19.2, pp. 487–525. MR: 2188134 (cit. on p. 1018).

Xiuxiong Chen, Simon K. Donaldson, and Song Sun (2015a). "Kähler-Einstein metrics on Fano manifolds. I: Approximation of metrics with cone singularities". *J. Amer. Math. Soc.* 28.1, pp. 183–197. MR: 3264766 (cit. on pp. 1012, 1017, 1023).

– (2015b). "Kähler-Einstein metrics on Fano manifolds. II: Limits with cone angle less than 2π". *J. Amer. Math. Soc.* 28.1, pp. 199–234. MR: 3264767 (cit. on pp. 1012, 1017, 1023).

– (2015c). "Kähler-Einstein metrics on Fano manifolds. III". *J. Amer. Math. Soc.* 28 (cit. on pp. 1012, 1017, 1023).

Xiuxiong Chen and W. Y. He (2008). "On the Calabi flow". *Amer. J. Math.* 130.2, pp. 539–570. MR: 2405167 (cit. on p. 1026).

Xiuxiong Chen and Song Sun (2014). "Calabi flow, geodesic rays, and uniqueness of constant scalar curvature Kähler metrics". *Ann. of Math. (2)* 180.2, pp. 407–454. MR: 3224716 (cit. on p. 1027).

Xiuxiong Chen, Song Sun, and Bing Wang (Aug. 2015). "Kähler-Ricci flow, Kähler-Einstein metric, and K-stability". arXiv: 1508.04397 (cit. on pp. 1012, 1025, 1026).

Xiuxiong Chen and G. Tian (2008). "Geometry of Kähler metrics and foliations by holomorphic discs". *Publ. Math. Inst. Hautes Études Sci.* 107, pp. 1–107. MR: 2434691 (cit. on p. 1013).

Xiuxiong Chen and Bing Wang (May 2014). "Space of Ricci flows (II)". arXiv: 1405.6797 (cit. on pp. 1016, 1025).

Xiuxiong Chen and Brian Weber (2011). "Moduli spaces of critical Riemannian metrics with $L^{\frac{n}{2}}$ norm curvature bounds". *Adv. Math.* 226.2, pp. 1307–1330. MR: 2737786 (cit. on p. 1018).

Xuemiao Chen and Song Sun (July 2017). "Singularities of Hermitian–Yang–Mills connections and the Harder-Narasimhan-Seshadri filtration". arXiv: 1707.08314 (cit. on p. 1021).

Craig van Coevering (2011). "Examples of asymptotically conical Ricci-flat Kähler manifolds". *Math. Z.* 267.1-2, pp. 465–496. MR: 2772262 (cit. on p. 1019).

Tristan C. Collins and Gábor Székelyhidi (Apr. 2012). "K-Semistability for irregular Sasakian manifolds". arXiv: 1204.2230 (cit. on p. 1020).

Tristan C. Collins and Valentino Tosatti (2015). "Kähler currents and null loci". *Invent. Math.* 202.3, pp. 1167–1198. MR: 3425388 (cit. on p. 1017).

Ronan J. Conlon and Hans-Joachim Hein (2015). "Asymptotically conical Calabi-Yau metrics on quasi-projective varieties". *Geom. Funct. Anal.* 25.2, pp. 517–552. MR: 3334234 (cit. on p. 1019).

Ved Datar and Gábor Székelyhidi (2016). "Kähler-Einstein metrics along the smooth continuity method". *Geom. Funct. Anal.* 26.4, pp. 975–1010. MR: 3558304 (cit. on p. 1012).

Ruadhaí Dervan and Gábor Székelyhidi (Dec. 2016). "The Kähler-Ricci flow and optimal degenerations". arXiv: 1612.07299 (cit. on p. 1025).

Simon K. Donaldson (2001). "Scalar curvature and projective embeddings. I". *J. Differential Geom.* 59.3, pp. 479–522. MR: 1916953 (cit. on p. 1015).

– (2004). "Conjectures in Kähler geometry". In: *Strings and geometry*. Vol. 3. Clay Math. Proc. Amer. Math. Soc., Providence, RI, pp. 71–78. MR: 2103718 (cit. on p. 1027).

– (2015). "Algebraic families of constant scalar curvature Kähler metrics". In: *Surveys in differential geometry 2014. Regularity and evolution of nonlinear equations*. Vol. 19. Surv. Differ. Geom. Int. Press, Somerville, MA, pp. 111–137. MR: 3381498 (cit. on p. 1024).

– (2018). "Some recent developments in Kähler geometry and exceptional holonomy". In: *Proc. of the International Congress of Mathematicians, Rio de Janeiro 2018*. Ed. by Boyan Sirakov, Paulo Ney de Souza, and Marcelo Viana. In this volume. World Scientific (cit. on p. 1013).

Simon K. Donaldson and Song Sun (2014). "Gromov-Hausdorff limits of Kähler manifolds and algebraic geometry". *Acta Math.* 213.1, pp. 63–106. MR: 3261011 (cit. on pp. 1014, 1016, 1017, 1020).

– (2017). "Gromov-Hausdorff limits of Kähler manifolds and algebraic geometry, II". *J. Differential Geom.* 107.2, pp. 327–371. MR: 3707646 (cit. on pp. 1018–1020).

Philippe Eyssidieux, Vincent Guedj, and Ahmed Zeriahi (2009). "Singular Kähler–Einstein metrics". *J. Amer. Math. Soc.* 22.3, pp. 607–639. MR: 2505296 (cit. on p. 1014).

Kento Fujita (Aug. 2015). "Optimal bounds for the volumes of Kähler-Einstein Fano manifolds". arXiv: 1508.04578 (cit. on pp. 1021, 1023).

Weiyong He (2016). "Kähler-Ricci soliton and H-functional". *Asian J. Math.* 20.4, pp. 645–663. MR: 3570456 (cit. on p. 1025).

Hans-Joachim Hein and Song Sun (2017). "Calabi-Yau manifolds with isolated conical singularities". *Publ. Math. Inst. Hautes Études Sci.* 126, pp. 73–130. arXiv: 1607.02940. MR: 3735865 (cit. on p. 1021).
C. LeBrun and S. R. Simanca (1994). "Extremal Kähler metrics and complex deformation theory". *Geom. Funct. Anal.* 4.3, pp. 298–336. MR: 1274118 (cit. on p. 1013).
Chi Li (Nov. 2015). "Minimizing normalized volumes of valuations". arXiv: 1511.08164 (cit. on pp. 1019, 1021).
Chi Li and Yuchen Liu (Feb. 2016). "Kähler-Einstein metrics and volume minimization". arXiv: 1602.05094 (cit. on p. 1021).
Chi Li, Xiaowei Wang, and Chenyang Xu (Nov. 2014). "Degeneration of Fano Kahler-Einstein manifolds". arXiv: 1411.0761 (cit. on p. 1024).
– (Feb. 2015). "Quasi-projectivity of the moduli space of smooth Kahler-Einstein Fano manifolds". arXiv: 1502.06532 (cit. on p. 1024).
Chi Li and Chenyang Xu (July 2017). "Stability of Valuations: Higher Rational Rank". arXiv: 1707.05561 (cit. on p. 1021).
Haozhao Li, Bing Wang, and Kai Zheng (Jan. 2015). "Regularity scales and convergence of the Calabi flow". arXiv: 1501.01851 (cit. on p. 1027).
Gang Liu (2016). "Gromov-Hausdorff limits of Kähler manifolds and the finite generation conjecture". *Ann. of Math. (2)* 184.3, pp. 775–815. MR: 3549623 (cit. on p. 1017).
Yuchen Liu (Sept. 2016). "The volume of singular Kähler-Einstein Fano varieties". arXiv: 1605.01034 (cit. on p. 1023).
Yuchen Liu and Chenyang Xu (June 2017). "K-stability of cubic threefolds". arXiv: 1706.01933 (cit. on p. 1023).
Toshiki Mabuchi and Shigeru Mukai (1993). "Stability and Einstein-Kähler metric of a quartic del Pezzo surface". In: *Einstein metrics and Yang-Mills connections (Sanda, 1990)*. Vol. 145. Lecture Notes in Pure and Appl. Math. Dekker, New York, pp. 133–160. MR: 1215285 (cit. on p. 1022).
Dario Martelli, James Sparks, and Shing-Tung Yau (2008). "Sasaki-Einstein manifolds and volume minimisation". *Comm. Math. Phys.* 280.3, pp. 611–673. MR: 2399609 (cit. on p. 1020).
Shigeru Mukai (1995). "New developments in Fano manifold theory related to the vector bundle method and moduli problems". *Sūgaku* 47.2, pp. 125–144. MR: 1364825 (cit. on p. 1022).
Yuji Odaka (Nov. 2012a). "On the moduli of Kahler-Einstein Fano manifolds". arXiv: 1211.4833 (cit. on p. 1024).
– (2012b). "The Calabi conjecture and K-stability". *Int. Math. Res. Not. IMRN* 10, pp. 2272–2288. MR: 2923166 (cit. on p. 1024).
– (2013). "The GIT stability of polarized varieties via discrepancy". *Ann. of Math. (2)* 177.2, pp. 645–661. MR: 3010808 (cit. on p. 1024).

– (2015). "Compact moduli spaces of Kähler-Einstein Fano varieties". *Publ. Res. Inst. Math. Sci.* 51.3, pp. 549–565. arXiv: 1412.2972. MR: 3395458 (cit. on p. 1024).

Yuji Odaka, Cristiano Spotti, and Song Sun (2016). "Compact moduli spaces of del Pezzo surfaces and Kähler-Einstein metrics". *J. Differential Geom.* 102.1, pp. 127–172. MR: 3447088 (cit. on pp. 1022, 1023).

Xiaochun Rong and Yuguang Zhang (2011). "Continuity of extremal transitions and flops for Calabi-Yau manifolds". *J. Differential Geom.* 89.2. Appendix B by Mark Gross, pp. 233–269. MR: 2863918 (cit. on p. 1017).

Jian Song (June 2017). "Degeneration of Kahler-Einstein manifolds of negative scalar curvature". arXiv: 1706.01518 (cit. on pp. 1017, 1024).

C. Spotti (n.d.). "Kähler-Einstein metrics via moduli continuity method" (cit. on p. 1022).

Cristiano Spotti and Song Sun (Apr. 2017). "Explicit Gromov-Hausdorff compactifications of moduli spaces of Kähler-Einstein Fano manifolds". arXiv: 1705.00377 (cit. on p. 1023).

Cristiano Spotti, Song Sun, and Chengjian Yao (2016). "Existence and deformations of Kähler-Einstein metrics on smoothable $\mathbb{Q}$-Fano varieties". *Duke Math. J.* 165.16, pp. 3043–3083. MR: 3566198 (cit. on p. 1024).

Jacopo Stoppa (2009). "K-stability of constant scalar curvature Kähler manifolds". *Adv. Math.* 221.4, pp. 1397–1408. MR: 2518643 (cit. on p. 1012).

Jeffrey Streets (2014). "Long time existence of minimizing movement solutions of Calabi flow". *Adv. Math.* 259, pp. 688–729. MR: 3197669 (cit. on p. 1027).

Gábor Székelyhidi (2009). "The Calabi functional on a ruled surface". *Ann. Sci. Éc. Norm. Supér. (4)* 42.5, pp. 837–856. MR: 2571959 (cit. on p. 1026).

G. Tian (1990a). "Kähler-Einstein metrics on algebraic manifolds". In: *Proceedings of the ICM 1990, Kyoto*. Ed. by 587–598. Vol. 1. Springer-Verlag (cit. on pp. 1015, 1016).

– (1990b). "On Calabi's conjecture for complex surfaces with positive first Chern class". *Invent. Math.* 101.1, pp. 101–172. MR: 1055713 (cit. on pp. 1014, 1016).

Gang Tian (1997). "Kähler-Einstein metrics with positive scalar curvature". *Invent. Math.* 130.1, pp. 1–37. MR: 1471884 (cit. on p. 1012).

Gang Tian and Jeff Viaclovsky (2008). "Volume growth, curvature decay, and critical metrics". *Comment. Math. Helv.* 83.4, pp. 889–911. MR: 2442967 (cit. on p. 1018).

Gang Tian and Shing-Tung Yau (1987). "Kähler-Einstein metrics on complex surfaces with $C_1 > 0$". *Comm. Math. Phys.* 112.1, pp. 175–203. MR: 904143 (cit. on p. 1022).

Valentino Tosatti (2015). "Families of Calabi-Yau manifolds and canonical singularities". *Int. Math. Res. Not. IMRN* 20, pp. 10586–10594. MR: 3455875 (cit. on p. 1017).

Shing-Tung Yau (1978). "On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I". *Comm. Pure Appl. Math.* 31.3, pp. 339–411. MR: 480350 (cit. on p. 1012).

Steve Zelditch (1998). “Szegő kernels and a theorem of Tian”. *Internat. Math. Res. Notices* 6, pp. 317–331. MR: 1616718 (cit. on p. 1015).

Received 2017-11-30.

Song Sun (孙崧)
song.sun@stonybrook.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1031–1058)

AN INVITATION TO HIGHER TEICHMÜLLER THEORY

Anna Wienhard

Abstract

Riemann surfaces are of fundamental importance in many areas of mathematics and theoretical physics. The study of the moduli space of Riemann surfaces of a fixed topological type is intimately related to the study of the Teichmüller space of that surface, together with the action of the mapping class group. Classical Teichmüller theory has many facets and involves the interplay of various methods from geometry, analysis, dynamics and algebraic geometry. In recent years, higher Teichmüller theory emerged as a new field in mathematics. It builds as well on a combination of methods from different areas of mathematics. The goal of my talk is to invite the reader to get to know and to get involved into higher Teichmüller theory by describing some of its many facets.

1 Introduction

Riemann surfaces are of fundamental importance in many areas of mathematics and theoretical physics. The study of the moduli space of Riemann surfaces of a fixed topological type is intimately related to the study of the Teichmüller space of that surface, together with the action of the mapping class group. Classical Teichmüller theory has many facets and involves the interplay of various methods from geometry, analysis, dynamics and algebraic geometry. In recent years, higher Teichmüller theory emerged as a new field in mathematics. It builds as well on a combination of methods from different areas of mathematics. The goal of this article is to invite the reader to get to know and to get involved into higher Teichmüller theory by describing some of its many facets. Along the way we point to open questions, and formulate some conjectures and task for the future. We will

The author acknowledges support by the National Science Foundation under agreement DMS-1536017, by the Sloan Foundation, by the Deutsche Forschungsgemeinschaft, by the European Research Council under ERC-Consolidator grant 614733, and by the Klaus-Tschira-Foundation.

MSC2010: primary 22E40; secondary 57M50, 13F60, 53D30, 57N16, 57S30.

Keywords: Higher Teichmüller theory, Total positivity, surface group representations, discrete subgroups of Lie groups, cluster algebras, geodesic flow, mapping class group.

not be able to discuss every aspect of higher Teichmüller theory, and will be very brief on most of them. In particular we will not touch upon universal or infinite higher Teichmüller spaces Labourie [2007a] and N. Hitchin [2016], or algebraic structures developed in Labourie [2017b] and Sun [2017a,b].

Higher Teichmüller theory is concerned with the study of representations of fundamental groups of oriented surface S of negative Euler characteristic into simple real Lie groups G of higher rank. The diversity of the methods involved is due partly to the one-to-one correspondence between representations, flat bundles, and Higgs bundles given by non-abelian Hodge theory Simpson [1991], which was established in work of Donaldson [1987, 1985], N. J. Hitchin [1987], Corlette [1988], and Simpson [1992].

We will introduce higher Teichmüller spaces below as special subsets of the representation variety $\mathrm{Hom}(\pi_1(S), G))/G$, namely as connected components consisting entirely of discrete and faithful representations. This is however a definition which only arose a posteriori. The first family of higher Teichmüller spaces, the Hitchin components, has been introduced by N. J. Hitchin [1992] using the theory of Higgs bundles. That they are higher Teichmüller spaces in the sense of our definition was in general only proven ten years later by Labourie [2006] and independently Fock and Goncharov [2006] through the study of the space of positive decorated local systems or positive representations. The second family of higher Teichmüller spaces, the spaces of maximal representations, was defined completely independently as the level set of a characteristic number on the representation variety, and its property of being a higher Teichmüller space in the sense of our definition was shown by Burger, Iozzi, and Wienhard [2003], motivated by previous work of W. M. Goldman [1988] in the context of classical Teichmüller space. The results of N. J. Hitchin [1992], Labourie [2006], Fock and Goncharov [2006] and Burger, Iozzi, and Wienhard [2003] arose completely independently, from different points of view and using very different methods. Only when comparing them it become apparent that the three spaces, Hitchin components, spaces of positive representations, and spaces of maximal representations, have many similarities and provide examples of a new phenomenon. Now we consider them as two families of what we call higher Teichmüller spaces. As the reader will see, we are still exploring the similarities and differences of these two families. It is interesting to note that the interplay between geometric and dynamical methods for representations of finitely generated groups and the more analytic and algebro-geometric methods from the theory of Higgs bundles are at the heart of several recent advances in our understanding of higher Teichmüller spaces.

Many questions in higher Teichmüller theory are motivated by the things we know about classical Teichmüller space, its properties and interesting geometric and dynamical structures it carries. However, there are also several new features that only arise for higher Teichmüller spaces and are not present in classical Teichmüller theory, see for example Section 4, Section 10 and Section 12. Higher Teichmüller theory is a very young and active field of mathematics. It is shaped by young mathematicians. There are still many open questions and unchartered territory to explore. We therefore hope that many young (and older) mathematicians will accept this invitation and contribute to the field in the future.

2 Classical Teichmüller space

Let S be a closed connected oriented topological surface of negative Euler characteristic $\chi(S) = 2 - 2g < 0$, where g is the genus of S. The Teichmüller space $\mathcal{T}(S)$ of S is the space of marked conformal classes of Riemannian metrics on S. It has been well studied using the theory of quasi-conformal maps as well as methods from hyperbolic geometry. By the uniformization theorem, there is a unique hyperbolic, i.e. constant curvature -1, metric in each conformal class. This identifies $\mathcal{T}(S)$ with the moduli space of marked hyperbolic structures. A marked hyperbolic structure is a pair (X, f_X), where X is a hyperbolic surface and $f_X : S \to X$ is an orientation preserving homeomorphism. Two marked hyperbolic structures (X, f_X) and (Y, f_Y) are equivalent if there exists an isometry $g : X \to Y$ such that $g \circ f_X$ is isotopic to f_Y. The mapping class group of S acts naturally on $T(S)$ by changing the marking. This action is properly discontinuous, and the quotient of $\mathcal{T}(S)$ by this action is the moduli space $\mathfrak{M}(S)$ of Riemann surfaces of topological type given by S. Teichmüller space is homeomorphic to $\mathbb{R}^{6g-6}$ and the universal cover of $\mathfrak{M}(S)$.

Higher Teichmüller theory builds on an algebraic realization of Teichmüller space. The universal cover $\tilde{X}$ of the hyperbolic surface X naturally identifies with the hyperbolic plane $\mathbb{H}^2$, and the fundamental group $\pi_1(X)$ acts as group of deck transformations by isometries on $\tilde{X} \cong \mathbb{H}^2$. Thus, upon fixing a base point, the marking induces a group homomorphism $(f_X)_* : \pi_1(S) \to \pi_1(X) < \mathrm{Isom}^+(\mathbb{H}^2) \cong \mathrm{PSL}(2, \mathbf{R})$, which is called the holonomy. Associating to a marked hyperbolic structure its holonomy gives a well defined injective map

$$\mathrm{hol} : \mathcal{T}(S) \to \mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2, \mathbf{R}))/\mathrm{PSL}(2, \mathbf{R}).$$

The representation variety $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2, \mathbf{R}))/\mathrm{PSL}(2, \mathbf{R})$ is the space of all group homomorphisms of $\pi_1(S)$ into $\mathrm{PSL}(2, \mathbf{R})$, up to conjugation by $\mathrm{PSL}(2, \mathbf{R})$. It carries a natural topology (induced from the topology of $\mathrm{PSL}(2, \mathbf{R})$). Teichmüller spce $\mathcal{T}(S)$ is a connected component of the representation variety $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2, \mathbf{R}))/\mathrm{PSL}(2, \mathbf{R})$. It is one of the two connected components, which consist entirely of discrete and faithful representations of $\pi_1(S)$ into $\mathrm{PSL}(2, \mathbf{R})$. The other such component is $\mathcal{T}(\overline{S})$, where $\overline{S}$ is the surface with the opposite orientation.

Remark 1. From this point of, as set of discrete and faithful representation, $\mathcal{T}(S)$ was first studied by Fricke. Historically it would thus be more appropriate to call it Fricke space and the generalizations higher Fricke spaces, but it seems hard to change a name that is well established.

Classical Teichmüller space has many interesting properties and carries additional structure. It is a Kähler manifold which admits several Riemannian and non-Riemannian metrics, has nice explicit parametrizations, and carries interesting flows and dynamical systems. We will not be able to recall most of these interesting properties, but will come back to a few of them in the sequel.

3 What is higher Teichmüller theory?

We might interpret what higher Teichmüller theory is in a narrow or a broader sense. In a very broad sense it is the study of classes of representations of finitely generated groups into Lie groups of higher rank with particularly nice geometric and dynamical behaviour. In the narrow sense one could characterize it as the study of higher Teichmüller spaces as we define them below. In this article we restrict most of our discussion to this narrow interpretation. In the broad sense it is touched upon also in the contributions of Kassel [n.d.] and Potrie [n.d.].

Teichmüller space is a connected component of the representation variety $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2,\mathbf{R}))/\mathrm{PSL}(2,\mathbf{R})$ - this is where higher Teichmüller theory takes it starting point. Instead of focussing on group homomorphisms of $\pi_1(S)$ into $\mathrm{PSL}(2,\mathbf{R})$, we replace $\mathrm{PSL}(2,\mathbf{R})$ by a simple Lie group G of higher rank (this is what the higher refers to), such as $\mathrm{PSL}(n,\mathbf{R})$, $n \geq 3$ or $\mathrm{Sp}(2n,\mathbf{R})$, $n \geq 2$, and consider the representation variety $\mathrm{Hom}(\pi_1(S), G)/G$. We make the following definition:

Definition 2. A higher Teichmüller space is a subset of $\mathrm{Hom}(\pi_1(S), G)/G$, which is a union of connected components that consist entirely of discrete and faithful representations.

Note that as soon as G is not locally isomorphic to $\mathrm{PSL}(2,\mathbf{R})$, the group $\pi_1(S)$ is not isomorphic to a lattice in G. Therefore the set of discrete and faithful representations is only a closed subset of $\mathrm{Hom}(\pi_1(S), G)/G$. It is thus not clear that higher Teichmüller spaces exist at all, and in fact they will only exist for special families of Lie groups G. In particular, when G is a simply connected complex Lie group, the representation variety $\mathrm{Hom}(\pi_1(S), G)/G$ is irreducible as an algebraic variety, and hence connected, and there cannot be any connected component consisting entirely of discrete and faithful representations. There are two known families of higher Teichmüller spaces, Hitchin components and spaces of maximal representations. They have been discovered from very different points of view and by very different methods. It then became clear that they share many properties, in particular the property requested in Definition 2. We describe a common underlying characterization, which also suggests the existence of two further families of higher Teichmüller spaces in Section 7.

Hitchin components $\mathcal{T}_H(S, G)$ are defined when G is a split real simple Lie group, the space of maximal representations $\mathcal{T}_{max}(S, G)$ is defined when G is a non-compact simple Lie group of Hermitian type. In the case when $G = \mathrm{PSL}(2, \mathbf{R})$, the Hitchin component and the space of maximal representations agree and coincide with Teichmüller space $\mathcal{T}(S)$. For other groups G not locally isomorphic to $\mathrm{PSL}(2, \mathbf{R})$, which are at the same time split and of Hermitian type, i.e. $\mathrm{Sp}(2n, \mathbf{R})$ or $\mathrm{SO}(2, 3)$, there is a proper inclusion $\mathcal{T}_H(S, G) \subset \mathcal{T}_{max}(S, G)$.

Remark 3. We assume that S is a closed surface.There is a related theory for surfaces with punctures or boundary components. However, in this case the corresponding subset of the representation variety is not a union of connected components. We comment on the situation for surfaces with punctures in Section 7.

We shortly review the definitions of Hitchin components and maximal representations. For more details and further properties we refer the reader to the survey Burger, Iozzi, and Wienhard [2014].

3.1 Hitchin components. Hitchin components are defined when G is a split real simple Lie group. Any split real simple Lie group G contains a three-dimensional principal subgroup, i.e. an embedding $\iota_p : \mathrm{SL}(2, \mathbf{R}) \to G$, which is unique up to conjugation. For the classical Lie groups $\mathrm{SL}(n, \mathbf{R})$, $\mathrm{Sp}(2n, \mathbf{R})$, and $\mathrm{SO}(n, n+1)$ this is just the irreducible representation of $\mathrm{SL}(2, \mathbf{R})$ in the appropriate dimension. Precomposing ι_p with a discrete embedding of $\pi_1(S)$ into $\mathrm{SL}(2, \mathbf{R})$ we obtain a representation $\rho_p : \pi_1(S) \to G$, which we call a principal Fuchsian representation.

Definition 4. The Hitchin component $\mathcal{T}_H(S, G)$ is the connected component of $\mathrm{Hom}(\pi_1(S), G)/G$ containing a principal Fuchsian representation $\rho_p : \pi_1(S) \to G$.

Remark 5. Note that we are a bit sloppy in our terminology, e.g. when $G = \mathrm{PSL}(3, \mathbf{R})$ there are 2 connected components in $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(3, \mathbf{R}))/\mathrm{PSL}(3, \mathbf{R})$ which consists of discrete and faithful representation which preserve the orientation. We refer to each of them as the Hitchin component.

Hitchin showed, using methods from the theory of Higgs bundle that the Hitchin component is homeomorphic to a vector space of dimension $\dim(G)(2g-2)$. He conjectured that these components are geometrically significant and parametrize geometric structures. This was supported by one example. W. M. Goldman [1990] had investigated the spaces of convex real projective structures on S and shown that it is isomorphic to $\mathbf{R}^{16g-16}$

and that the holonomy of a convex real projective structure is in the Hitchin component. Soon afterwards Choi and W. M. Goldman [1993] proved that in fact $\mathcal{T}_H(S, \mathrm{PSL}(3, \mathbf{R}))$ parametrizes the space of convex real projective structures on S. It took another ten years before further progress was made, when Labourie [2006] introduced methods from dynamical systems to the study of representations in the Hitchin component and showed that for $G = \mathrm{PSL}(n, \mathbf{R}), \mathrm{PSp}(2n, \mathbf{R})$, and $\mathrm{PO}(n, n+1)$ representations in the Hitchin component are discrete and faithful. That representations in any Hitchin component are discrete and faithful follows from work of Fock and Goncharov [2006]. They investigated the space of positive representations in $\mathrm{Hom}(\pi_1(S), G)/G$, when G is a split real simple Lie group, and showed that it coincides with the Hitchin component (see Section 7).

3.2 Maximal representations. Maximal representations are defined when the simple Lie group G is of Hermitian type. They are singled out by a characteristic number, the Toledo number, which for $G = PSL(2, \mathbf{R})$ is just the Euler number. W. M. Goldman [1988] showed that the Euler number distinguishes the connected components of $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2, \mathbf{R}))/\mathrm{PSL}(2, \mathbf{R})$, and that Teichmüller space corresponds to the connected component formed by representation of Euler number $2g - 2$, which is the maximal value it can attain. In general, the Toledo number is bounded in terms of the Euler characteristic of S and the real rank of G, and constant on connected components of $\mathrm{Hom}(\pi_1(S), G)/G$. The space of maximal representations $\mathcal{T}_{max}(S, G)$ is the set of all representations for which the Toledo number assumes it maximal possible value. It is a union of connected components. Using methods from bounded cohomology, it was proven in Burger, Iozzi, and Wienhard [2003] that any maximal representation is faithful with discrete image.

Remark 6. There are two types of Hermitian Lie groups, those of tube type and those not of tube-type. Maximal representations into Lie groups that are not of tube type satisfy a rigidity theorem Toledo [1989], Hernàndez [1991], Burger, Iozzi, and Wienhard [2003], S. B. Bradlow, García-Prada, and Gothen [2003], and S. B. Bradlow, García-Prada, and Gothen [2006]: The image of a maximal representation is always contained in the stabilizer in G of a maximal subsymmetric space of tube type. This reduces the study of maximal representation essentially to the case when G is of tube type.

3.3 Anosov representations. Anosov representation are homomorphisms of finitely generated hyperbolic groups Γ into arbitrary reductive Lie groups G with special dynamical properties. They have been introduced by Labourie [2006] to investigate representations in the Hitchin component, and extended to hyperbolic groups in Guichard and Wienhard [2012]. The set of Anosov representations is an open subset of $\mathrm{Hom}(\Gamma, G)/G$, but in general not a union of connected components of $\mathrm{Hom}(\Gamma, G)/G$. Representations in the Hitchin component and maximal representations were the first examples of Anosov representations Labourie [2006], Burger, Iozzi, Labourie, and Wienhard [2005], Guichard and Wienhard [2012], and Burger, Iozzi, and Wienhard [n.d.]. We refer to Kassel's contribution Kassel [n.d.] for the definition, more details and more references on Anosov representations.

The following key properties of Hitchin representations and maximal representations follow from them being Anosov representations (with respect to certain parabolic subgroups).

1. Every representation in the Hitchin component and every maximal representation is discrete and faithful.

2. Let $\rho : \pi_1(S) \to G$ be a Hitchin representation, then there exists a ρ-equivariant continuous boundary map $\xi : S^1 \to G/B$ into the generalized flag variety G/B, where B is the Borel subgroup of G. The map sends distinct points in S^1 to transverse points in G/B.

3. Let $\rho : \pi_1(S) \to G$ be a maximal representation, then there exists a ρ-equivariant continuous boundary map $\xi : S^1 \to G/S$ into the generalized flag variety G/S, where S is a maximal parabolic subgroup of G which fixes a point in the Shilov boundary of the symmetric space $X = G/K$. The map sends distinct points in S^1 to transverse points in G/S.

4 Topology of the representation variety

For a connected Lie group G the obstruction to lifting a representation $\pi_1(S) \to G$ to the universal cover of G defines a characteristic invariant in $\mathrm{H}^2(S, \pi_1(G)) \cong \pi_1(G)$. For compact simple Lie groups Atiyah and Bott [1983] and complex simple Lie groups W. M. Goldman [1988] and J. Li [1993] the connected components of $\mathrm{Hom}(\pi_1(S), G)/G$ are in one to one correspondence with elements in $\pi_1(G)$. This does not hold anymore for real simple Lie groups in general. Of course, characteristic invariants in $\mathrm{H}^2(S, \pi_1(G))$ still distinguish some of the connected components, but they are not sufficient to distinguish all of them. N. J. Hitchin [1992] determined the number of connected components of $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(n, \mathbf{R}))/\mathrm{PSL}(n, \mathbf{R})$, and showed that Hitchin components have the same characteristic invariants as other components. The space of maximal representations, which is defined using characteristic invariants, in fact decomposes itself into several connected components, which hence cannot be distinguished by any characteristic invariant Gothen [2001] and S. B. Bradlow, García-Prada, and Gothen [2006].

A precise count of the number of connected components for several classical groups, and in particular for the connected components of the space of maximal representations has been given using the Morse theoretic methods on the moduli space of Higgs bundles Hitchin introduced A. G. Oliveira [2011], García-Prada and A. G. Oliveira [2014], S. B. Bradlow, García-Prada, and Gothen [2015], Gothen [2001], S. B. Bradlow, García-Prada, and Gothen [2006], García-Prada and Mundet i Riera [2004], and García-Prada, Gothen, and Mundet i Riera [2013]. The situation is particularly interesting for $G = \mathrm{Sp}(4, \mathbf{R})$ (and similarly for the locally isomorphic group $\mathrm{SO}^\circ(2,3)$) as there are $2g - 4$ connected components in which all representations are Zariski dense Guichard and Wienhard [2010] and S. B. Bradlow, García-Prada, and Gothen [2012]. Maximal representations in these components cannot be obtained by deforming an appropriate Fuchsian representation $\rho : \pi_1(S) \to \mathrm{SL}(2, \mathbf{R}) \to \mathrm{Sp}(4, \mathbf{R})$. An explicit construction of representations in these exceptional components is given in Guichard and Wienhard [2010]. Any maximal representations into $\mathrm{Sp}(2n, \mathbf{R})$ with $n \geq 3$ on the other hand can be deformed either to a principal Fuchsian representation (if it is in a Hitchin component) or to a (twisted) diagonal Fuchsian representation $\rho : \pi_1(S) \to \mathrm{SL}(2, \mathbf{R}) \times \mathrm{O}(n) < \mathrm{Sp}(2n, \mathbf{R})$. In order to distinguish the connected components in the space of maximal representations, additional invariants are necessary. Such additional invariants have been defined on the one hand using methods from the theory of Higgs bundles Gothen [2001], S. B. Bradlow, García-Prada, and Gothen [2006], Collier [2017], Baraglia and Schaposnik [2017], and Aparicio-Arroyo, S. Bradlow, Collier, García-Prada, Gothen, and A. Oliveira [2018] and on the other hand using the Anosov property of representations Guichard and Wienhard [2010].

We shortly describe the additional invariants arising from the Anosov property. If a representations is Anosov with respect to a parabolic subgroup $P < G$, then the pull-back of the associated flat G-bundle to $T^1 S$ admits a reduction of the structure group to L, where L is the Levi subgroup of P. The characteristic invariants of this L-bundle provide additional invariants of the representation. These additional invariants can be used in particular to further distinguish connected components consisting entirely of Anosov representation. Note that representations can be Anosov with respect to different parabolic subgroups - each such parabolic subgroup gives rise to additional invariants. For the case of maximal representations into $\mathrm{Sp}(2n, \mathbf{R})$ it is shown in Guichard and Wienhard [ibid.] that these additional invariants in fact distinguish all connected components.

Conjecture 7. *Let G be a simple Lie group of higher rank. The connected components of $\mathrm{Hom}(\pi_1(S), G)/G$ can be distinguished by characteristic invariants and by additional invariants associated to unions of connected components consisting entirely of Anosov representations.*

Note that, since Anosov representations are discrete and injective, any connected component consisting entirely of Anosov representations also provides an example of a higher Teichmüller space. Thus Conjecture 7 implies in particular the following

Conjecture 8. *There are connected components of* $\mathrm{Hom}(\pi_1(S), G)/G$ *which are not distinguished by characteristic invariants, if and only if there exist higher Teichmüller spaces in* $\mathrm{Hom}(\pi_1(S), G)/G$

Combining Conjecture 8 with Conjecture 19 gives a precise list of groups (see Theorem 17) for which we expect additional connected components to exist. A particularly interesting case is $\mathrm{SO}(p,q)$, $p \neq q$. Here additional components and additional invariants have recently been found via Higgs bundle methods Collier [2017], Baraglia and Schaposnik [2017], and Aparicio-Arroyo, S. Bradlow, Collier, García-Prada, Gothen, and A. Oliveira [2018].

5 Geometric Structures

Classical Teichmüller space $\mathcal{T}(S)$ is not just a space of representations, but in fact a space of geometric structures: every representation is the holonomy of a hyperbolic structure on S. For higher Teichmüller spaces, such a geometric interpretation is less obvious. The quotient of the symmetric space Y associated to G by $\rho(\pi_1(S))$ is of infinite volume. In order to find geometric structures on compact manifolds associated, other constructions are needed.

For any representation $\rho : \pi_1(S) \to G$ in the Hitchin component or in the space of maximal representation, there is t a domain of discontinuity in a generalized flag variety $X = G/Q$, on which ρ acts cocompactly. The quotient is a compact manifold M with a locally homogeneous (G, X)-structure. This relies on the the construction of domains of discontinuity for Anosov representations given by Guichard and Wienhard [2012] and generalized by Kapovich, Leeb, and Porti [2018]. We do not describe this construction here in detail, but refer the reader to Kassel [n.d.], where locally homogeneous (G, X)-structures, Anosov representations and the construction of domains of discontinuity are discussed in more detail.

The construction of the domains of discontinuity, together with some topological considerations, allows one to deduce the general statement

Theorem 9. *Guichard and Wienhard [2012] For every split real simple Lie group* G *there exists a generalized flag variety* X *and a compact manifold* M *such that* $\mathcal{T}_H(S, G)$ *parametrizes a connected component of the deformation space of* (G, X)*-structures on* M*. For every Lie group of Hermitian type* G *there exists a generalized flag variety* X *and a compact manifold* M *such that for every connected component* C *of* $\mathcal{T}_{max}(S, G)$ *the following holds: A Galois cover of* C *parametrizes a connected component of the deformation space of* (G, X)*-structures on* M*.*

In particular, any Hitchin representation or maximal representation is essentially the holonomy of a (G, X)-structures on a compact manifold. It is however quite hard to get an explicit description of the deformation space of (G, X)-structures they parametrize. First it is nontrivial to determine the topology of the quotient manifold, and second it is rather difficult to give a synthetic description of the geometric properties which ensure that the holonomy representation of a (G, X) structure lies in the Hitchin component or in the space of maximal representations. In three cases we such a synthetic description: the Hitchin component for $\mathcal{T}_H(S, \mathrm{PSL}(3, \mathbf{R}))$, the Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(4, \mathbf{R}))$ and $\mathcal{T}_H(S, \mathrm{PSp}(4, \mathbf{R}))$, the space of maximal representations $\mathcal{T}_{max}(S, \mathrm{SO}^\circ(2, n))$.

Theorem 10. Choi and W. M. Goldman [1993] *A representation $\rho : \pi_1(S) \to \mathrm{PSL}(3, \mathbf{R})$ is in Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(3, \mathbf{R}))$ if and only if it is the holonomy representation of a convex real projective structures on S*

A convex real projective structure on S is a realization of S as the quotient of a convex domain $\Omega \subset \mathbf{RP}^2$ by a group $\Gamma < \mathrm{PSL}(3, \mathbf{R})$ of projective linear transformation preserving Ω. One aspect which makes this case very special is that the group $\mathrm{PSL}(3, \mathbf{R})$ acts as transformation group on the two-dimensional homogeneous space $\mathbf{RP}^2$. The subgroup $\rho(\pi_1(S))$ preserves a convex domain Ω and acts cocompactly on it. The quotient is a surface homeomorphic to S. For more general simple Lie groups of higher rank, there is no two-dimensional generalized flag variety on which they act, and so the quotient manifold M is higher dimensional.

Theorem 11. Guichard and Wienhard [2008] *The Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(4, \mathbf{R}))$ parametrizes the space of properly convex foliated projective structures on the unit tangent bundle of S. The Hitchin component $\mathcal{T}_H(S, \mathrm{PSp}(4, \mathbf{R}))$ parametrizes the space of properly convex foliated projective contact structures on the unit tangent bundle of S.*

Theorem 12. Collier, Tholozan, and Toulisse [2017] *The space of maximal representations $\mathcal{T}_{max}(S, \mathrm{SO}^\circ(2, n))$ parametrizes the space of fibered photon structures on $\mathrm{O}(n)/\mathrm{O}(n-2)$ bundles over S.*

Conjecture 13 (Guichard-Wienhard)**.** *Let $\rho : \pi_1(S) \to G$ be a representation in a higher Teichmüller space, then there exists a generalized flag variety X and compact fiber bundle $M \to S$, such that $\overline{\rho} : \pi_1(M) \to \pi_1(S) \to G$, where $\pi_1(M) \to \pi_1(S)$ is induced by the bundle map and $\pi_1(S) \to G$ is given by ρ, is the holonomy of a locally homogeneous (G, X)-structure on M.*

In fact, we expect, that for any cocompact domain of discontintuity which is constructed through a balanced thickening in the sense of Kapovich, Leeb, and Porti [2018], the quotient manifold is homeomorphic to a compact fiber bundle M over S. A related conjecture has been made by Dumas and Sanders [2017b, Conjecture 1.1] for deformations of

Hitchin representations in the complexification of G. They proved the conjecture in the case of $\mathrm{PSL}(3,\mathbf{C})$. Guichard and Wienhard [2011a] determine the topology of the quotient manifold for maximal representations and Hitchin representations into the symplectic group, Alessandrini and Q. Li [2018] prove the conjecture for Hitchin representations into $\mathrm{PSL}(n,\mathbf{R})$, and their deformations into $\mathrm{PSL}(n,\mathbf{C})$, and Alessandrini, Maloni, and Wienhard [n.d.] analyze the topology of quotient manifold for complex deformations of symplectic Hitchin representations.

It is very interesting to note that the recent advantages Alessandrini and Q. Li [2018] and Collier, Tholozan, and Toulisse [2017] on understanding the topology of the quotient manifolds rely on a finer analysis and description of the Higgs bundle associated to special representations in the Hitchin component or in the space of maximal representations. That the explicit description of the Higgs bundles can be used to endow the domain of discontinuity naturally with the structure of a fiber bundle was first described by Baraglia [2010] for $\mathfrak{T}_H(S,\mathrm{PSL}(4,\mathbf{R}))$, where he recovered the projective structures on the unit tangent bundle from Theorem 11.

6 Relation to the moduli space of Riemann surfaces

The outer automorphism group fo $\pi_1(S)$ is isomorphic to the mapping class group of S. It acts naturally on $\mathrm{Hom}(\pi_1(S),G)/G$. This action is properly discontinuous on higher Teichmüller spaces - in fact more generally on the set of Anosov representations Labourie [2008], Wienhard [2006], and Guichard and Wienhard [2012]. It is natural to ask about the relation between the quotient of higher Teichmüller spaces by this action and the moduli space of Riemann surfaces $\mathfrak{M}(S)$. For Hitchin components Labourie made a very precise conjecture, based on Hitchin's parametrization of the Hitchin component. We state the parametrization and Labourie's conjecture for $G = \mathrm{PSL}(n,\mathbf{R})$ to simplify notation. Hitchin introduced the Hitchin component in N. J. Hitchin [1992] using methods from the theory of Higgs bundles. This requires the choice of a conformal structure on S. N. J. Hitchin [ibid.] showed, using methods from the theory of Higgs bundles, that the Hitchin component is homeomorphic to a vector space. Namely, it is homeomorphic to the space of holomorphic differentials on S with respect to a chosen conformal structure, i.e. $\mathfrak{T}_H(S,\mathrm{PSL}(n,\mathbf{R})) \cong \sum_{i=2}^n H^0(S,K^i)$ This parametrization depends on the choice of a conformal structure and is not invariant under the mapping class group.

Conjecture 14. *Labourie [2008] The quotient of $\mathfrak{T}_H(S,\mathrm{PSL}(n,\mathbf{R}))$ by the mapping class group is a holomorphic vector bundle over $\mathfrak{M}(S)$, with fiber equal to $\sum_{i=3}^n H^0(S,K^i)$.*

This conjecture has been proven by Labourie [2007b] and Loftin [2001] for $G = \mathrm{PSL}(3, \mathbf{R})$ and by Labourie [2017a], using Higgs bundle methods, for all split real Lie groups of rank 2. It is open in all other cases.

For maximal representations we expect similarly to get a mapping class group invariant projection from $\mathfrak{T}_{max}(S, G)$ to $\mathfrak{T}(S)$, see Alessandrini and Collier [2017, Conjecture 10]. For $G = \mathrm{SO}(2, n)$ such a projection is constructed in Collier, Tholozan, and Toulisse [2017]. For G of rank 2 Alessandrini and Collier [2017] construct a mapping class group invariant complex structure on $\mathfrak{T}_{max}(S, G)$. For $G = \mathrm{Sp}(4, \mathbf{R})$ they show that the quotient of $\mathfrak{T}_{max}(S, G)$ by the mapping class group is a holomorphic vector bundle over $\mathfrak{M}(S)$, and describe in detail the fiber over a point, which is rather complicated since the space of maximal representations has nontrivial topology and singular points.

7 Positivity

Hitchin components and maximal representation were introduced and studied by very different methods. It turns out that they do not only share many properties, but also admit a common characterization in terms of positive structures on flag varieties. Only the flag varieties and notions of positivity in question are different for Hitchin components and maximal representations.

For Hitchin components we consider full flag varieties and Lusztig's total positivity Lusztig [1994]. For maximal representations the flag variety in question is the Shilov boundary of the symmetric space of G and positivity is given by the Maslov cocycle. In order to keep the description simple, we illustrate both notions in examples. We consider $G = \mathrm{SL}(n, \mathbf{R})$ for Hitchin components, and $G = \mathrm{Sp}(2n, \mathbf{R})$ for maximal representations.

The relevant flag variety for $\mathfrak{T}_H(S, \mathrm{SL}(n, \mathbf{R}))$ is the full flag variety

$$\mathfrak{F}(\mathbf{R}^n) := \{F = (F_1, F_2, \cdots, F_{n-1}) \mid F_i \subset \mathbf{R}^n, \dim(F_i) = i, \; F_i \subset F_{i+1}\}.$$

Two flags F, F' are said to be transverse if $F_i \cap F'_{n-i} = \{0\}$. We fix the standard basis $(e_1, \cdots, e_n)$ of $\mathbf{R}^n$. Let $F \in \mathfrak{F}$ be the flag with $F_i = \mathrm{span}(e_1, \cdots, e_i)$, and $E \in \mathfrak{F}$ the flag with $E_i = \mathrm{span}(e_n, \cdots, e_{n-i+1})$.

Any flag T transverse to F, is the image of E under a unique unipotent matrix u_T. The triple of flags (E, T, F) is said to be *positive* if and only if u_T is a totally positive unipotent matrix. Note that a unipotent (here lower triangular) matrix is totally positive if and only of every minor is positive, except those that have to be zero by the condition that the matrix is unipotent. Any two transverse flags (F_1, F_2) can be mapped to (E, F) by an element of $\mathrm{SL}(n, \mathbf{R})$ and we can extend the notion of positivity to any triple of pairwise transverse flags.

Theorem 15. *Fock and Goncharov [2006], Labourie [2006], and Guichard [2008] Let $\rho : \pi_1(S) \to \mathrm{SL}(n, \mathbf{R})$ be a representation. Then $\rho \in \mathcal{T}_H(S, \mathrm{SL}(n, \mathbf{R}))$ if and only if there exists a continuous ρ-equivariant map $\xi : S^1 \to \mathcal{F}(\mathbf{R}^n)$ which sends positive triples in S^1 to positive triples in $\mathcal{F}(\mathbf{R}^n)$.*

To describe the analogous characterization of maximal representations into $\mathrm{Sp}(2n, \mathbf{R})$ we consider $\mathbf{R}^{2n}$ with the standard symplectic form ω and let $\{e_1, \cdots, e_n, f_1, \cdots f_n\}$ be a symplectic basis. Then

$$\mathcal{L}(\mathbf{R}^{2n}) := \{L \subset \mathbf{R}^{2n} \mid \dim L = n,\ \omega|_{L \times L} = 0\}$$

is the space of Lagrangian subspaces, and two Lagrangians L and L' are transverse if $L \cap L' = \{0\}$.

Fix $L_E = \mathrm{span}(e_1, \cdots, e_n)$ and $L_F = \mathrm{span}(f_1, \cdots, f_n)$. Any Lagrangian $L_T \in \mathcal{L}$ transverse to L_F is the image of L_E under an element $v_T = \begin{pmatrix} Id_n & 0 \\ M_T & Id_n \end{pmatrix} \in V$, where M_T is a symmetric matrix.

The triple of Lagrangians (L_E, L_T, L_F) is said to be *positive* if and only if $M_T \in \mathrm{Pos}(n, \mathbf{R}) \subset \mathrm{Sym}(n, \mathbf{R})$ is positive definite. This is equivalent to the Maslov cocycle of (L_E, L_T, L_F) being n, which is the maximal value it can attain. The symplectic group $\mathrm{Sp}(2n, \mathbf{R})$ acts transitively on the space of pairs of transverse Lagrangians and we can extend the notion of positivity to any triple of pairwise transverse Lagrangian.

Theorem 16. *Burger, Iozzi, and Wienhard [2003] Let $\rho : \pi_1(S) \to \mathrm{Sp}(2n, \mathbf{R}$ be a representation. Then $\rho \in \mathcal{T}_{max}(S, \mathrm{Sp}(2n, \mathbf{R}))$ if and only if there exists a continuous ρ-equivariant map $\xi : S^1 \to \mathcal{L}(\mathbf{R}^{2n})$ which sends positive triples in S^1 to positive triples in $\mathcal{L}(\mathbf{R}^{2n}$.*

In Guichard and Wienhard [2011c,b] we introduce the notion of Θ-positivity. It generalizes Lusztig's total positivity, which is only defined for split real Lie groups, to arbitrary simple Lie groups. There are four families of Lie groups admitting a Θ-positive structure:

Theorem 17. *Guichard and Wienhard [2011c, Theorem 4.3.] A simple Lie group G admits a Θ-positive structure if and only if:*

1. *G is a split real form.*
2. *G is of Hermitian type of tube type.*
3. *G is locally isomorphic to $\mathrm{SO}(p, q)$, $p \neq q$,.*
4. *G is a real form of F_4, E_6, E_7, E_8, whose restricted root system is of type F_4.*

Θ refers to a subset of the simple roots. If G admits a Θ-positive structure, then there is positive semigroup $U_\Theta^> \subset U_\Theta < P_\Theta$ with which we can define the notion of positivity for a triple of pairwise transverse points in the generalized flag variety G/P_Θ as above. Here P_Θ is the parabolic group associated to the subset of simple roots Θ, and U_Θ is its unipotent radical.

Definition 18. A representation $\rho : \pi_1(S) \to G$ is said to be Θ-positive if there exists a continuous ρ-equivariant map $\xi : S^1 \to G/P_\Theta$ which sends positive triples in S^1 to positive triples in G/P_Θ.

Conjecture 19 (Guichard-Labourie-Wienhard)**.** *The set of Θ-positive representations $\rho : \pi_1(\Sigma_g) \to G$ is open and closed in* $\mathrm{Hom}(\pi_1(S), G)/G$. *In particular, Θ-positive representations form higher Teichmüller spaces.*

For more details on Θ-positivity and Θ-positive representations we refer the reader to Guichard and Wienhard [2011c] and the upcoming papers Guichard and Wienhard [2011b] and Guichard, Labourie, and Wienhard [2011], in which Conjecture 13 will be partly addressed. In particular we prove that Θ-positive representations are P_Θ-Anosov and form an open subset of $\mathrm{Hom}(\pi_1(S), G)/G$, and a closed set, at least in the subset of irreducible representations.

The existence of a Θ-positive structure provides a satisfying answer on when and why higher Teichmüller spaces exist, and we expect that the families of Lie groups listed in Theorem 17 are the only simple Lie groups for which higher Teichmüller spaces in $\mathrm{Hom}(\pi_1(S), G)/G$ exist. A particular interesting case is the family of Θ-positive representations for $G = \mathrm{SO}(p, q)$. Here the connected components have recently been determined with Higgs bundle methods, and several of them contain Θ-positive representations Collier [2017] and Aparicio-Arroyo, S. Bradlow, Collier, García-Prada, Gothen, and A. Oliveira [2018].

8 Coordinates and Cluster structures

Teichmüller space carries several nice sets of coordinates. The best known are Fenchel-Nielsen coordinates, which encode a hyperbolic structure by the length of and the twist around a set of $3g-3$ disjoint simple closed non-homotopic curves which give a decomposition of S into a union of $2g-2$ pair of pants. W. M. Goldman [1990] introduced Fenchel-Nielsen type coordinates on the Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(3, \mathbf{R})$. Here, there are two length and two twist coordinates associated to the curves of a pants decomposition, and in addition two coordinates which associated to each of the pairs of pants. This a new feature arises because a convex real projective structure on a pair of pants is not uniquely determined by the holonomies around the boundary. For maximal representations, Fenchel-Nielsen type coordinates were constructed in Strubel [2015].

It is often easier to describe coordinates in the situation when the surface is not closed, but has at least one puncture. In this case one can consider decorated flat bundles (or decorated representations), which is a flat bundle or a representation together with additional information around the puncture. Fixing an ideal triangulation, i.e. a triangulation where all the vertices are punctures, this additional information can be used to define coordinates. Examples of this are Thurston shear coordinates or Penner coordinates for decorated Teichmüller space. In the context of higher Teichmüller spaces, for decorated representations into split real Lie groups Fock and Goncharov [2006] introduced two sets of coordinates, so called $\mathcal{X}$- coordinates, which generalize Thurston shear coordinates, and $\mathcal{A}$-coordinates, which generalize Penner coordinates. They show that when performing a flip of the triangulation (changing the diagonal in a quadrilateral formed by two adjacent triangles), the change of coordinates is given by a positive rational function. As a consequence, the set of decorated representations where all coordinates are positive, is independent of the triangulation. In fact, Fock and Goncharov prove that this set of positive representations is precisely the set of positive representations in the sense of Section 7, where the notion of positivity stems from Lusztig's positivity. In the case when $G = \mathrm{PSL}(n, \mathbf{R})$, the Fock-Goncharov coordinates admit a particularly nice geometric description based on triple ratios and cross ratios. In particular there is a close relation between the coordinates and cluster structures, which received a lot of attention. The change of coordinates associated to a flip of the triangulations is given by a sequence of cluster mutations. This has since been generalized to other classical groups in Le [n.d.(b)], see also Le [n.d.(a)] and Goncharov and Shen [2018] for general split real Lie groups. Related coordinates have been defined by Gaiotto, Moore and Neitzke, using the theory of spectral networks Gaiotto, Moore, and Neitzke [2013, 2014]. For an interpretation of the Weil-Petersson form in terms of cluster algebras see Gekhtman, Shapiro, and Vainshtein [2005].

Inspired by Fock-Goncharov coordinates for surfaces with punctures, Bonahon and Dreyer defined coordinates on the Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$ and showed that $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$ is real analytically homeomorphic to the interior of a convex polygon of dimension $(n^2-1)(2g-2)$ Bonahon and Dreyer [2014, 2017]. These coordinates are associated to a maximal lamination of the surface S and generalize Thurston's shear coordinates of closed surfaces. A special case for such a maximal lamination is an ideal triangulation of S which is subordinate to a pair of pants decomposition, i.e. the lamination consists of $3g-3$ disjoint simple closed non-homotopic curves which give a pair of pants decomposition, and three curves in each pair of pants, that cut the pair of pants into two ideal triangles. In this case, Zhang [2015a] provided a reparametrization of the Bonahon-Dreyer coordinates, which give a genuine generalization of Fenchel-Nielsen type coordinates for the Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$, see also Bonahon and Kim [2016] for a direct comparision with Goldman coordinates when $n = 3$.

In forthcoming work Alessandrini, Guichard, Rogozinnikov, and Wienhard [2017] we introduce $\mathcal{X}$-type and $\mathcal{A}$-type coordinates for decorated maximal representations of the fundamental group of a punctured surface into the symplectic group $\mathrm{Sp}(2n, \mathbf{R})$. These coordinates have the feature of behaving like the coordinates for $G = \mathrm{PSL}(2, \mathbf{R})$ but with values in the space of positive definite symmetric matrices $\mathrm{Pos}(n, \mathbf{R})$. In particular, even though they are noncommutative, they exhibit a cluster structure. This structure is similar to the noncommutative cluster structure considered by Berenstein and Retakh [2015], except for a difference in some signs.

It would be interesting to develop similar coordinates for Θ-positive representations, in particular for those into $\mathrm{SO}(p, q)$, and to investigate their properties. The properties of the Θ-positive structure suggests that in this case the cluster-like structure would combine noncommutative and commutative aspects.

Task 20. *Develop $\mathcal{X}$-type and $\mathcal{A}$-type coordinates for decorated Θ-positive representations into $\mathrm{SO}(p, q)$. Analyze their cluster-like structures.*

9 Symplectic geometry and dynamics

For any reductive Lie group, the representation variety of a closed surface $\mathrm{Hom}(\pi_1(S), G)/G$ is a symplectic manifold W. M. Goldman [1984]. On Teichmüller space this symplectic structure interacts nicely with Fenchel-Nielsen coordinates. The length and twist coordinates give global Darboux coordinates: the length coordinate associated to a simple closed curve in a pair of pants decomposition is symplectically dual to the twist coordinate associated to this curve, and the symplectic form can be expressed by Wolpert's formula as $\omega = \sum_{i=1}^{3g-3} dl_i \wedge d\tau_i$, where l_i is the length coordinate and τ_i is the twist coordinate Wolpert [1983]. The twist flows associated to a simple closed curve c on S is the flow given by cutting S along c and continuously twisting around this curve before gluing the surface back together. It is the Hamiltonian flow associated to the length coordinate defined by c. The twist flows associated to the $3g - 3$ simple closed curves in a pants decomposition on S commute. This gives Teichmüller space the structure of a complete integrable system. For more general reductive groups, the Hamiltonian flows associated to length functions on $\mathrm{Hom}(\pi_1(S), G)/G$ have been studied by W. M. Goldman [1986].

In Sun and Zhang [2017] provide a new approach to compute the Goldman symplectic form on $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$. This, in conjunction with a companion article by Sun, Wienhard, and Zhang [2017], gives rise to several nice statements. Given maximal lamination with finitely many leaves (and some additional topological data) we construct in Sun, Wienhard, and Zhang [ibid.] new families of flows on $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$. These flows give a trivialization of the Hitchin component, which is shown to be symplectic in Sun and Zhang [2017]. Consequently the flows are all Hamiltonian flows and provide $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R})$ the structure of a completely integrable system.

Task 21. *The mapping class group* $\mathrm{Mod}(S)$ *acts naturally on the space of maximal laminations and the additional topological data, so that the symplectic trivializations of* $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$ *induce representations* $\pi_n : \mathrm{Mod}(S) \to \mathrm{Sp}(\mathbf{R}^{(n^2-1)(2g-2)})$. *Analyze these representations.*

A special situation arises when the maximal lamination is an ideal triangulation subordinate to a pants decomposition. In this situation we slightly modify the Bonahon-Dreyer coordinates, to get global Darboux coordinates on $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$ which consist of $(3g-3)(n-1)$ length coordinates, $(3g-3)(n-1)$ twist coordinates associated to the simple closed curves in the pants decomposition, and $2 \times \frac{(n-1)(n-2)}{2}$ coordinates for each pair of pants. The twist flows are the Hamiltonian flows associated to the length functions, and for each pair of pants we introduce $\frac{(n-1)(n-2)}{2}$ new flows, which we call eruption flows. Their Hamiltonian functions are rather complicated. Nevertheless, the twist flows and the eruption flows pairwise commute, providing a half dimensional subspace of commuting flows. In the case of $\mathrm{PSL}(3, \mathbf{R})$ the eruption flow has been defined in Wienhard and Zhang [2018], where it admits a very geometric description.

Classical Teichmüller space does not only admit twist flows, but carries several natural flows, for example earthquake flows, which extend twist flows, geodesic flows with respect to the Weil-Petersson metric or the Teichmüller metric or, and even an $\mathrm{SL}(2, \mathbf{R})$-action. None of this has yet been explored for higher Teichmüller spaces. A new approach for lifting Teichmüller dynamics to representation varieties for general Lie group G has recently been described by Forni and W. Goldman [2017].

10 Geodesic flows and entropy

Representations in higher Teichmüller space, and more generally Anosov representations, are strongly linked to dynamics on the surface S. Any such representation gives rise to a Hölder reparametrization of the geodesic flow on S and the representation can essentially be reconstructed from the periods of this reparametrized geodesic flow. This dynamical point of view has been first observed by Labourie [2006] and applied by Sambarino [2014a] and has been key in several interesting developments.

Using the thermodynamical formalism Bridgeman, Canary, Labourie, and Sambarino [2015] define the pressure metric on the Hitchin component $\mathcal{T}_H(S, \mathrm{PSL}(n, \mathbf{R}))$ and more generally spaces of Anosov representations. The pressure metric restricts to a multiple of the Weil-Petersson metric on the subset of principal Fuchsian representations. In the case of $\mathrm{PSL}(3, \mathbf{R})$ metrics on the space of convex real projective structures have also been constructed in Darvishzadeh and W. M. Goldman [1996] and Q. Li [2016].

Other important quantities that have been investigated using this dynamical viewpoint are the critical exponent and the topological entropy of Hitchin representations, which are related to counting orbit points on the symmetric space Sambarino [2015, 2014b] and Pollicott and Sharp [2014]. Here completely new features arise that are not present in classical Teichmüller space. On Teichmüller space both quantities are constant, but on Hitchin components these functions vary and provide information about the geometry of the representations. There are sequences of representations along which the entropy goes to zero Zhang [2015b,a]. The entropy is in fact bounded above. Tholozan for $n = 3$ Tholozan [2017], and Potrie and Sambarino in general, establish an entropy rigidity theorem: A representation saturates the upper bound for the entropy if and only if it is a principal Fuchsian representation Sambarino [2016] and Potrie and Sambarino [2017]. This has consequences for the volume of the minimal surface in the symmetric space associated to the representation. A key aspect in the work of Potrie and Sambarino has been the regularity of the map $\xi : S^1 \to G/B$ of a Hitchin representation.

For maximal representations which are not in the Hitchin component much less is known. One obstacle is the missing regularity of the boundary map $\xi : S^1 \to G/S$, which has rectifiable image, but is in general not smooth. Glorieux and Monclair [n.d.] study the entropy of Anti-de-Sitter Quasi-Fuchsian representations $\pi_1(S) \to \mathrm{SO}(2, n)$, some of their methods methods might also be useful to investigate maximal representation $\pi_1(S) \to \mathrm{SO}(2, n)$.

Task 22. *Investigate the topological entropy of maximal representations. Find and characterize sequences along which the entropy goes to zero. Find bounds for the topological entropy and geometrically characterize the representations that saturate these bounds.*

A lot of the geometry of Teichmüller space can be recovered from geodesic currents associated to the representations and from their intersection Bonahon [1988]. Recently Martone and Zhang [n.d.] have associated geodesic currents to positively ratioed representations, a class that includes Hitchin representations and maximal representations but should include also Θ-positive representations. Bridgeman, Canary, Labourie, and Sambarino [2018] define the Liouville current for a Hitchin representation which they use to construct the Liouville pressure metric. From the intersection of the geodesic currents one can recover the periods of the reparametrization of the geodesic flow and the periods of crossratios associated to the representation Labourie [2008] and Hartnick and Strubel [2012]. These geodesic currents and the corresponding crossratio functions also play an important role for the next topic we discuss.

11 Compactifications

Classical Teichmüller space admits various non-homeomorphic compactifications. One such compactification is the marked length spectrum compactification. The marked length spectrum of a representation $\rho : \pi_1(S) \to \mathrm{PSL}(2,\mathbf{R})$ associates to any conjugacy class in $[\gamma]$ in $\pi_1(S)$ the translation length of the element $\rho(\gamma)$ in $\mathbb{H}^2$. It is a basic result that for Teichmüller space the map $\Phi : \mathcal{T}(S) \to \mathbf{P}(\mathbf{R}^{\mathcal{C}})$ provides an embedding. The closure of $\Phi(\mathcal{T}(S))$ is the marked length spectrum compactification. It is homeomorphic to the Thurston compactification of $\mathcal{T}(S)$ by the space of projectivized measured laminations. This compactification has been reconstructed using geodesic currents by Bonahon [1988].

The marked length spectrum compactification has been generalized to compactifications of spaces of representations of finitely generated groups into reductive Lie groups by Parreau [2012], where the translation length in $\mathbb{H}^2$ is replaced by the vector valued translation length in the symmetric space associated to G. This can be applied to Hitchin components and spaces of maximal representations to provide marked length spectrum compactifications. The investigation of the fine structure of these compactifications has just begun. A key ingredient are generalizations of the Collar Lemma from hyperbolic geometry to Hitchin representations Lee and Zhang [2017] and to maximal representations Burger and M. B. Pozzetti [2017]. See also Labourie and McShane [2009], Vlamis and Yarmola [2017], and Fanoni and B. Pozzetti [n.d.] for generalizations of cross-ratio identities in the context of higher Teichmüller spaces. Burger, Iozzi, Parreau, and B. Pozzetti [2017] establish a new decomposition theorem for geodesic currents, which will play a crucial role in their program to understand the marked length spectrum compactification of maximal representations.

Compactifications of the space of positive representations (when S has punctures) have been constructed using explicit parametrizations and the theory of tropicalizations Fock and Goncharov [2006], Alessandrini [2008], and Le [2016].

In all these constructions it is a challenge to give a geometric interpretation of points in the boundary of the compactification. The most natural is in terms of actions on $\mathbf{R}$-buildings Parreau [2012], Le [2016], and Burger and M. B. Pozzetti [2017]. This naturally generalizes the description of boundary points of Teichmüller space by actions on $\mathbf{R}$-trees. However, Thurston's compactification gives an interpretation of boundary points by measured laminations on S. It would be interesting to get a description of boundary points of Hitchin components or spaces of maximal representations in terms of geometric objects on S that generalize measured laminations, and to relate the compactifications to degenerations of the geometric structures associated to Hitchin components and maximal representations in low dimensions (see Section 5). For convex real projective structures such an interpretation of boundary points in terms of a mixed structure consisting of a measured lamination and a HEX-metric has been announced by Cooper, Delp, D. Long, and Thistlethwaite [n.d.].

12 Arithmetics

Discrete Zariski dense subgroups of semisimple Lie groups have been studied for a long time. In recent years there has been a revived interest from number theory in discrete Zariski dense subgroups, which are contained in arithmetic lattices without being lattices themselves. Such groups have been coined thin groups Sarnak [2014]. Hitchin representations and maximal representations, and more generally Anosov representations, provide many examples of discrete Zariski dense subgroups of higher rank Lie groups which are not lattices. They are not necessarily contained in (arithmetic) lattices. However, there are many Hitchin representations and maximal representations that are even integral, i.e. up to conjugation contained in the integral points of the group G. This is a new feature that arises for higher Teichmüller spaces and is not present in classical Teichmüller space. The group $\mathrm{PSL}(2,\mathbf{Z})$ is a non-uniform lattice in $\mathrm{PSL}(2,\mathbf{R})$, consequently there is no discrete embedding of the fundamental group $\pi_1(S)$ of a closed oriented surface into $\mathrm{PSL}(2,\mathbf{R})$ which takes values in the integer points $\mathrm{PSL}(2,\mathbf{Z})$.

For $n = 3$ the first examples of integral Hitchin representations were given by Vinberg and Kac [1967], infinite families were constructed by Long, Reid, and Thistlethwaite with an explicit description using triangle groups.

Theorem 23. *D. D. Long, Reid, and Thistlethwaite [2011] Let $\Gamma = \langle a, b \mid a^3 = b^3 = (ab)^4 = 1\rangle$ be the (3,3,4) triangle group. There is an explicit polynomial map $\phi : \mathbf{R} \to$ $\mathrm{Hom}(\Gamma, \mathrm{PSL}(3,\mathbf{R}))$ whose image lies in the Hitchin component. For all $t \in \mathbf{Z}$, the image of $\phi(t)$ give a Zariski-dense subgroup of $\mathrm{PSL}(3,\mathbf{Z})$. The representations $\phi(t)$, $t \in Z$ are pairwise not conjugate in $\mathrm{PGL}(3,\mathbf{R})$.*

Since Γ contains subgroups of finite index which are isomorphic to the fundamental group $\pi_1(S)$ of a closed oriented surface S, Theorem 23 gives rise to infinitely many, non conjugate integral representations in $\mathcal{T}_H(\pi_1(S), \mathrm{PSL}(n,\mathbf{R}))$. The representations lie on different mapping class group orbits.

In unpublished work with Burger and Labourie we use bending to show the following

Theorem 24. *For $n \geq 5$ and odd there are infinitely many pairwise non conjugate integral representations in the Hitchin component $\mathcal{T}_H(\pi_1(S), \mathrm{PSL}(n,\mathbf{R}))$. These representations lie on different mapping class group orbits.*

Task 25. *Develop tools to count integral representations in $\mathcal{T}_H(\pi_1(S), \mathrm{PSL}(n,\mathbf{R}))$ modulo the action of the mapping class group. Investigate the counting functions and their asymptotics.*

A first step to start counting is to find appropriate height functions on $\mathfrak{T}_H(\pi_1(S), \mathrm{PSL}(n, \mathbf{R}))$ such that there are only finitely many integral representations of finite height. A height function, inspired by Thurston's asymmetric metric on Teichmüller space has been proposed by Burger and Labourie.

For more general constructions of surface subgroups in lattices of Lie groups, following the construction of Kahn and Markovic surface subgroups in three-manifold groups Kahn and Markovic [2012] we refer to work of Hamenstädt and Kahn [2017], and forthcoming work of Labourie, Kahn, and Mozes [2017]. Examples of integral maximal representations are constructed in Toledo [1987].

13 Complex Analytic Theory

In classical Teichmüller theory complex analytic methods and the theory of quasi-conformal mappings play a crucial role. These aspects are so far largely absent from higher Teichmüller theory. Dumas and Sanders started exploring the complex analytic aspects of discrete subgroups of complex Lie groups of higher rank in Dumas and Sanders [2017b]. They investigate in particular deformations of Hitchin representations and maximal representations in the complexifications of G, and establish important properties of the complex compact manifolds M that arise as quotients of domains of discontinuity of these representations (see Section 5). In a forthcoming paper Dumas and Sanders [2017a] the complex deformation theory of these representations will be investigated further.

14 Higher dimensional higher Teichmüller spaces

Fundamental groups of surfaces are not the only finitely generated groups for which there are special connected components in the representation variety $\mathrm{Hom}(\pi_1(S), G)/G$, which consist entirely of discrete and faithful representations. This phenomenon also arises for fundamental groups of higher dimensional manifolds, and even for more general finitely generated hyperbolic groups. The main examples are convex divisible representations, which have been introduced and studied by Benoist in a series of papers, starting with Benoist [2004]. They are generalizations of convex real projective structures on surfaces, and exist in any dimension.

Let N be a compact hyperbolic manifold of dimension n and $\pi_1(N)$ its fundamental group. A representation $\rho : \pi_1(N) \to \mathrm{PGL}(n+1, \mathbf{R})$ is convex divisible if there exists a strictly $\rho(\pi_1(N))$- invariant convex domain in $\mathbf{RP}^n$, on which $\rho(\pi_1(N))$ acts cocompactly.

Theorem 26. *Benoist [2005] The set of convex divisible representations is a union of connected components of* $\mathrm{Hom}(\pi_1(N), \mathrm{PGL}(n+1, \mathbf{R}))/\mathrm{PGL}(n+1, \mathbf{R})$ *consisting entirely of discrete and faithful representations.*

Barbot [2015] showed that a similar phenomenon arise for representations $\rho : \pi_1(N) \to \mathrm{SO}(2, n)$ that are Anti-de-Sitter Quasi-Fuchsian representation. In fact these are precisely Anosov representations of $\pi_1(N)$ into $\mathrm{SO}(2, n)$ with respect to the parabolic subgroup that stabilizes an isotropic line Barbot and Mérigot [2012].

Theorem 27 (Barbot [2015])**.** *The set of Quasi Fuchsian AdS representations in* $\mathrm{Hom}(\pi_1(N), \mathrm{SO}(2, n))/\mathrm{SO}(2, n)$ *is a connected component consisting entirely of discrete and faithful representations.*

In view of Definition 2 we might call these connected components of the representation variety $\mathrm{Hom}(\pi_1(N), G)/G$ containing only discrete and faithful representations higher dimensional higher Teichmüller spaces.

When N is of dimension two, the notion of Θ-positivity gives us a conjectural criterion why and when higher Teichmüller spaces exist. It would be interesting to find a unifying principle behind the existence of such special connected components in $\mathrm{Hom}(\pi_1(N), G)/G$.

Task 28. *Find the underlying principle for the existence of connected components of the representation variety* $\mathrm{Hom}(\pi_1(N), G)/G$ *which consist entirely of discrete and faithful representations.*

A first test case could be to analyze deformations of the representation $\rho_0 : \pi_1(N) \to \mathrm{SO}(1, n) \to \mathrm{SO}(k, n)$, or more generally deformations of representations $\rho_0 : \pi_1(N) \to \mathrm{SO}(1, n) \to G$, where the centralizer of $\mathrm{SO}(1, n)$ in G is compact and contained in the maximal compact subgroup of the Levi group of a parabolic subgroup containing the parabolic subgroup defined by $\mathrm{SO}(1, n)$. We expect all such deformations to be discrete and faithful.

References

Daniele Alessandrini (2008). "Tropicalization of group representations". *Algebr. Geom. Topol.* 8.1, pp. 279–307. MR: 2443230 (cit. on p. 1049).

Daniele Alessandrini and Brian Collier (2017). "The geometry of maximal components of the $\mathrm{PSp}(4, \mathbb{R})$-character variety". arXiv: 1708.05361 (cit. on p. 1042).

Daniele Alessandrini, Olivier Guichard, Evgenii Rogozinnikov, and Anna Wienhard (2017). "Coordinates for maximal representations". in preparation (cit. on p. 1046).

Daniele Alessandrini and Qiongling Li (2018). "Projective structures with (quasi)-Hitchin holonomy". Preprint (cit. on p. 1041).

Daniele Alessandrini, Sara Maloni, and Anna Wienhard (n.d.). "The geometry of quasi-Hitchin symplectic representations". in preparation (cit. on p. 1041).

Marta Aparicio-Arroyo, Steven Bradlow, Brian Collier, Oscar García-Prada, Peter B. Gothen, and André Oliveira (2018). "SO(p, q)-Higgs bundles and higher Teichmüller components". arXiv: 1802.08093 (cit. on pp. 1038, 1039, 1044).

Michael F. Atiyah and Raoul Bott (1983). "The Yang-Mills equations over Riemann surfaces". *Philos. Trans. Roy. Soc. London Ser. A* 308.1505, pp. 523–615 (cit. on p. 1037).

David Baraglia (2010). "G2 geometry and integrable systems". PhD thesis. arXiv: 1002.1767 (cit. on p. 1041).

David Baraglia and Laura Schaposnik (2017). "Cayley and Langlands type correspondences for orthogonal Higgs bundles". arXiv: 1708.08828 (cit. on pp. 1038, 1039).

Thierry Barbot (2015). "Deformations of Fuchsian AdS representations are quasi-Fuchsian". *J. Differential Geom.* 101.1, pp. 1–46. MR: 3356068 (cit. on p. 1052).

Thierry Barbot and Quentin Mérigot (2012). "Anosov AdS representations are quasi-Fuchsian". *Groups Geom. Dyn.* 6.3, pp. 441–483 (cit. on p. 1052).

Yves Benoist (2004). "Convexes divisibles. I". In: *Algebraic groups and arithmetic*. Mumbai: Tata Inst. Fund. Res., pp. 339–374 (cit. on p. 1051).

– (2005). "Convexes divisibles. III". *Ann. Sci. École Norm. Sup. (4)* 38.5, pp. 793–832 (cit. on p. 1052).

Arkady Berenstein and Vladimir Retakh (2015). "Noncommutative marked surfaces". To appear in Advances in Mathematics. arXiv: 1510.02628 (cit. on p. 1046).

Francis Bonahon (1988). "The geometry of Teichmüller space via geodesic currents". *Invent. Math.* 92.1, pp. 139–162 (cit. on pp. 1048, 1049).

Francis Bonahon and Guillaume Dreyer (2014). "Parameterizing Hitchin components". *Duke Math. J.* 163.15, pp. 2935–2975. MR: 3285861 (cit. on p. 1045).

– (2017). "Hitchin characters and geodesic laminations". *Acta Math.* 218.2, pp. 201–295. MR: 3733100 (cit. on p. 1045).

Francis Bonahon and Inkang Kim (2016). "The Goldman and Fock-Goncharov coordinates for convex projective structures on surfaces". arXiv: 1607.03650 (cit. on p. 1046).

Steven B. Bradlow, Oscar García-Prada, and Peter B. Gothen (2003). "Surface group representations and $\mathrm{U}(p,q)$-Higgs bundles". *J. Differential Geom.* 64.1, pp. 111–170 (cit. on p. 1036).

Steven B. Bradlow, Oscar García-Prada, and Peter B. Gothen (2006). "Maximal surface group representations in isometry groups of classical Hermitian symmetric spaces". *Geom. Dedicata* 122, pp. 185–213. MR: 2295550 (cit. on pp. 1036–1038).

Steven B. Bradlow, Oscar García-Prada, and Peter B. Gothen (2012). "Deformations of maximal representations in $\mathrm{Sp}(4,\mathbb{R})$". *Q. J. Math.* 63.4, pp. 795–843 (cit. on p. 1038).

Steven B. Bradlow, Oscar García-Prada, and Peter B. Gothen (2015). "Higgs bundles for the non-compact dual of the special orthogonal group". *Geom. Dedicata* 175, pp. 1–48. MR: 3323627 (cit. on p. 1038).

Martin Bridgeman, Richard Canary, François Labourie, and Andres Sambarino (2015). "The pressure metric for Anosov representations". *Geom. Funct. Anal.* 25.4, pp. 1089–1179. MR: 3385630 (cit. on p. 1047).

– (2018). "Simple root flows for Hitchin representations". *Geom. Dedicata* 192, pp. 57–86. MR: 3749423 (cit. on p. 1048).

Marc Burger, Alessandra Iozzi, François Labourie, and Anna Wienhard (2005). "Maximal representations of surface groups: Symplectic Anosov structures". *Pure Appl. Math. Q.* 1.Special Issue: In memory of Armand Borel, pp. 543–590 (cit. on p. 1037).

Marc Burger, Alessandra Iozzi, Anne Parreau, and Beatrice Pozzetti (2017). "A structure theorem for geodesic currents and length spectrum compactifications". arXiv: 1710.07060 (cit. on p. 1049).

Marc Burger, Alessandra Iozzi, and Anna Wienhard (n.d.). "Maximal Representations and Anosov Structures". draft available on request (cit. on p. 1037).

– (2003). "Surface group representations with maximal Toledo invariant". *C. R. Math. Acad. Sci. Paris* 336.5, pp. 387–390 (cit. on pp. 1032, 1036, 1043).

– (2014). "Higher Teichmüller spaces: from $\mathrm{SL}(2,\mathbb{R})$ to other Lie groups". In: *Handbook of Teichmüller theory. Vol. IV*. Vol. 19. IRMA Lect. Math. Theor. Phys. Eur. Math. Soc., Zürich, pp. 539–618. MR: 3289711 (cit. on p. 1035).

Marc Burger and Maria Beatrice Pozzetti (2017). "Maximal representations, non-Archimedean Siegel spaces, and buildings". *Geom. Topol.* 21.6, pp. 3539–3599. MR: 3692972 (cit. on p. 1049).

Suhyoung Choi and William M. Goldman (1993). "Convex real projective structures on closed surfaces are closed". *Proc. Amer. Math. Soc.* 118.2, pp. 657–661 (cit. on pp. 1036, 1040).

Brian Collier (2017). "$\mathrm{SO}(n,n+1)$-surface group representations and their Higgs bundles". arXiv: 1710.01287 (cit. on pp. 1038, 1039, 1044).

Brian Collier, Nicolas Tholozan, and Jérémy Toulisse (2017). "The geometry of maximal representations of surface groups into $\mathrm{SO}(2,n)$". arXiv: 1702.08799 (cit. on pp. 1040–1042).

Daryl Cooper, Kelly Delp, Darren Long, and Morwen Thistlethwaite (n.d.). Talks and private communication (cit. on p. 1050).

Kevin Corlette (1988). "Flat G-bundles with canonical metrics". *J. Differential Geom.* 28.3, pp. 361–382 (cit. on p. 1032).

Mehdi-Reza Darvishzadeh and William M. Goldman (1996). "Deformation spaces of convex real-projective structures and hyperbolic affine structures". *J. Korean Math. Soc.* 33.3, pp. 625–639. MR: 1419757 (cit. on p. 1048).

Simon K. Donaldson (1985). “Anti self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles”. *Proc. London Math. Soc. (3)* 50.1, pp. 1–26 (cit. on p. 1032).

– (1987). “Twisted harmonic maps and the self-duality equations”. *Proc. London Math. Soc. (3)* 55.1, pp. 127–131 (cit. on p. 1032).

David Dumas and Andrew Sanders (2017a). in preparation (cit. on p. 1051).

– (2017b). “Geometry of compact complex manifolds associated to generalized quasi-Fuchsian representations”. arXiv: 1704.01091 (cit. on pp. 1040, 1051).

Federica Fanoni and Beatrice Pozzetti (n.d.). “Basmajian-type inequalities for maximal representations”. arXiv: 1611.00286 (cit. on p. 1049).

Vladimir Fock and Alexander Goncharov (2006). “Moduli spaces of local systems and higher Teichmüller theory”. *Publ. Math. Inst. Hautes Études Sci.* 103, pp. 1–211 (cit. on pp. 1032, 1036, 1043, 1045, 1049).

Giovanni Forni and William Goldman (2017). “Mixing Flows on Moduli Spaces of Flat Bundles over Surfaces”. arXiv: 1707.03099 (cit. on p. 1047).

Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke (2013). “Spectral networks”. *Ann. Henri Poincaré* 14.7, pp. 1643–1731. MR: 3115984 (cit. on p. 1045).

– (2014). “Spectral networks and snakes”. *Ann. Henri Poincaré* 15.1, pp. 61–141. MR: 3147409 (cit. on p. 1045).

Oscar García-Prada, Peter B. Gothen, and Ignasi Mundet i Riera (2013). “Higgs bundles and surface group representations in the real symplectic group”. *J. Topol.* 6.1, pp. 64–118 (cit. on p. 1038).

Oscar García-Prada and Ignasi Mundet i Riera (2004). “Representations of the fundamental group of a closed oriented surface in $\mathrm{Sp}(4, \mathbb{R})$”. *Topology* 43.4, pp. 831–855 (cit. on p. 1038).

Oscar García-Prada and André G. Oliveira (2014). “Connectedness of the moduli of $\mathrm{Sp}(2p, 2q)$-Higgs bundles”. *Q. J. Math.* 65.3, pp. 931–956. MR: 3261975 (cit. on p. 1038).

Michael Gekhtman, Michael Shapiro, and Alek Vainshtein (2005). “Cluster algebras and Weil-Petersson forms”. *Duke Math. J.* 127.2, pp. 291–311. MR: 2130414 (cit. on p. 1045).

Olivier Glorieux and Daniel Monclair (n.d.). “Critical exponent and Hausdorff dimension for quasi-Fuchsian AdS manifolds”. arXiv: 1606.05512 (cit. on p. 1048).

William M. Goldman (1984). “The symplectic nature of fundamental groups of surfaces”. *Adv. in Math.* 54.2, pp. 200–225 (cit. on p. 1046).

– (1986). “Invariant functions on Lie groups and Hamiltonian flows of surface group representations”. *Invent. Math.* 85.2, pp. 263–302. MR: 846929 (cit. on p. 1046).

– (1988). “Topological components of spaces of representations”. *Invent. Math.* 93.3, pp. 557–607 (cit. on pp. 1032, 1036, 1037).

William M. Goldman (1990). “Convex real projective structures on compact surfaces”. *J. Differential Geom.* 31.3, pp. 791–845 (cit. on pp. 1035, 1044).

Alexander Goncharov and Linhui Shen (2018). “Donaldson–Thomas transformations of moduli spaces of G-local systems”. *Adv. Math.* 327, pp. 225–348. MR: 3761995 (cit. on p. 1045).

Peter B. Gothen (2001). “Components of spaces of representations and stable triples”. *Topology* 40.4, pp. 823–850 (cit. on pp. 1037, 1038).

Olivier Guichard (2008). “Composantes de Hitchin et représentations hyperconvexes de groupes de surface”. *J. Differential Geom.* 80.3, pp. 391–431 (cit. on p. 1043).

Olivier Guichard, François Labourie, and Anna Wienhard (2011). “Positive representations”. in preparation (cit. on p. 1044).

Olivier Guichard and Anna Wienhard (2008). “Convex foliated projective structures and the Hitchin component for $\mathrm{PSL}_4(\mathbf{R})$”. *Duke Math. J.* 144.3, pp. 381–445 (cit. on p. 1040).

– (2010). “Topological invariants of Anosov representations”. *J. Topol.* 3.3, pp. 578–642 (cit. on p. 1038).

– (2011a). “Domains of discontinuity for maximal symplectic representations”. in preparation (cit. on p. 1041).

– (2011b). “Generalizing Lusztig positivity”. in preparation (cit. on pp. 1043, 1044).

– (2011c). “Positivity and higher Teichmüller theory”. to appear in Proceedings of the 7th ECM. arXiv: 1802.02833 (cit. on pp. 1043, 1044).

– (2012). “Anosov representations: domains of discontinuity and applications”. *Invent. Math.* 190, pp. 357–438 (cit. on pp. 1036, 1037, 1039, 1041).

Ursula Hamenstädt and Jeremy Kahn (2017). “Geometry of compact complex manifolds associated to generalized quasi-Fuchsian representations”. in preparation (cit. on p. 1051).

Tobias Hartnick and Tobias Strubel (2012). “Cross ratios, translation lengths and maximal representations”. *Geom. Dedicata* 161, pp. 285–322 (cit. on p. 1048).

Luis Hernàndez (1991). “Maximal representations of surface groups in bounded symmetric domains”. *Trans. Amer. Math. Soc.* 324, pp. 405–420 (cit. on p. 1036).

Nigel Hitchin (2016). “Higgs bundles and diffeomorphism groups”. In: *Surveys in differential geometry 2016. Advances in geometry and mathematical physics*. Vol. 21. Surv. Differ. Geom. Int. Press, Somerville, MA, pp. 139–163. MR: 3525096 (cit. on p. 1032).

Nigel J. Hitchin (1987). “The self-duality equations on a Riemann surface”. *Proc. London Math. Soc. (3)* 55.1, pp. 59–126 (cit. on p. 1032).

– (1992). “Lie groups and Teichmüller space”. *Topology* 31.3, pp. 449–473 (cit. on pp. 1032, 1037, 1041).

Jeremy Kahn and Vladimir Markovic (2012). "Immersing almost geodesic surfaces in a closed hyperbolic three manifold". *Ann. of Math. (2)* 175.3, pp. 1127–1190. MR: 2912704 (cit. on p. 1051).

Michael Kapovich, Bernhard Leeb, and Joan Porti (2018). "Dynamics on flag manifolds: domains of proper discontinuity and cocompactness". *Geom. Topol.* 22.1, pp. 157–234. MR: 3720343 (cit. on pp. 1039, 1040).

Fanny Kassel (n.d.). "GEOMETRIC STRUCTURES AND REPRESENTATIONS OF DISCRETE GROUPS". to appear in Proceedings of the ICM 2018 (cit. on pp. 1034, 1037, 1039).

François Labourie (2006). "Anosov flows, surface groups and curves in projective space". *Invent. Math.* 165, pp. 51–114 (cit. on pp. 1032, 1036, 1037, 1043, 1047).

François Labourie (2007a). "Cross ratios, surface groups, $\mathrm{PSL}(n, \mathbf{R})$ and diffeomorphisms of the circle". *Publ. Math. Inst. Hautes Études Sci.* 106, pp. 139–213. MR: 2373231 (cit. on p. 1032).

– (2007b). "Flat projective structures on surfaces and cubic holomorphic differentials". *Pure Appl. Math. Q.* 3.4. , Special Issue: In honor of Grigory Margulis. Part 1, pp. 1057–1099. MR: 2402597 (cit. on p. 1042).

François Labourie (2008). "Cross ratios, Anosov representations and the energy functional on Teichmüller space". *Ann. Sci. Éc. Norm. Supér. (4)* 41.3, pp. 437–469 (cit. on pp. 1041, 1048).

François Labourie (2017a). "Cyclic surfaces and Hitchin components in rank 2". *Ann. of Math. (2)* 185.1, pp. 1–58. MR: 3583351 (cit. on p. 1042).

François Labourie (2017b). "Goldman Algebra, Opers and the Swapping Algebra". arXiv: 1212.5015 (cit. on p. 1032).

François Labourie, Jeremy Kahn, and Shahar Mozes (2017). "Geometry of compact complex manifolds associated to generalized quasi-Fuchsian representations". in preparation (cit. on p. 1051).

François Labourie and Gregory McShane (2009). "Cross ratios and identities for higher Teichmüller-Thurston theory". *Duke Math. J.* 149.2, pp. 279–345. MR: 2541705 (cit. on p. 1049).

Ian Le (n.d.[a]). "An Approach to Cluster Structures on Moduli of Local Systems for General Groups". arXiv: 1606.00961 (cit. on p. 1045).

– (n.d.[b]). "Cluster Structures on Higher Teichmüller Spaces for Classical Groups". arXiv: 1603.03523 (cit. on p. 1045).

– (2016). "Higher laminations and affine buildings". *Geom. Topol.* 20.3, pp. 1673–1735. MR: 3523066 (cit. on p. 1049).

Gye-Seon Lee and Tengren Zhang (2017). "Collar lemma for Hitchin representations". *Geom. Topol.* 21.4, pp. 2243–2280. MR: 3654108 (cit. on p. 1049).

Jun Li (1993). “The space of surface group representations”. *Manuscripta Math.* 78.3, pp. 223–243 (cit. on p. 1037).

Qiongling Li (2016). “Teichmüller space is totally geodesic in Goldman space”. *Asian J. Math.* 20.1, pp. 21–46. MR: 3460757 (cit. on p. 1048).

John C. Loftin (2001). “Affine spheres and convex $\mathbb{RP}^n$-manifolds”. *Amer. J. Math.* 123.2, pp. 255–274. MR: 1828223 (cit. on p. 1042).

Darren D. Long, Alan W. Reid, and Morwen Thistlethwaite (2011). “Zariski dense surface subgroups in $\mathrm{SL}(3, \mathbf{Z})$”. *Geom. Topol.* 15.1, pp. 1–9. MR: 2764111 (cit. on p. 1050).

George Lusztig (1994). “Total positivity in reductive groups”. In: *Lie theory and geometry*. Vol. 123. Progr. Math. Boston, MA: Birkhäuser Boston, pp. 531–568 (cit. on p. 1042).

Guiseppe Martone and Tengren Zhang (n.d.). “Positively ratioed representations”. To appear in Comm. Math. Helv. arXiv: 1609.01245 (cit. on p. 1048).

André Gama Oliveira (2011). “Representations of surface groups in the projective general linear group”. *Internat. J. Math.* 22.2, pp. 223–279. MR: 2782688 (cit. on p. 1038).

Anne Parreau (2012). “Compactification d’espaces de représentations de groupes de type fini”. *Math. Z.* 272.1-2, pp. 51–86. MR: 2968214 (cit. on p. 1049).

Mark Pollicott and Richard Sharp (2014). “Length asymptotics in higher Teichmüller theory”. *Proc. Amer. Math. Soc.* 142.1, pp. 101–112. MR: 3119185 (cit. on p. 1048).

Rafael Potrie (n.d.). “Robust dynamics, invariant structures and topological classification”. To appear in Proceedings of the ICM 2018. arXiv: 1802.05291 (cit. on p. 1034).

Rafael Potrie and Andrés Sambarino (2017). “Eigenvalues and entropy of a Hitchin representation”. *Invent. Math.* 209.3, pp. 885–925. MR: 3681396 (cit. on p. 1048).

A. Sambarino (2014a). “Hyperconvex representations and exponential growth”. *Ergodic Theory Dynam. Systems* 34.3, pp. 986–1010. MR: 3199802 (cit. on p. 1047).

– (2016). “On entropy, regularity and rigidity for convex representations of hyperbolic manifolds”. *Math. Ann.* 364.1-2, pp. 453–483. MR: 3451394 (cit. on p. 1048).

Andrés Sambarino (2014b). “Quantitative properties of convex representations”. *Comment. Math. Helv.* 89.2, pp. 443–488. MR: 3229035 (cit. on p. 1048).

– (2015). “The orbital counting problem for hyperconvex representations”. *Ann. Inst. Fourier (Grenoble)* 65.4, pp. 1755–1797. MR: 3449196 (cit. on p. 1048).

Peter Sarnak (2014). “Notes on thin matrix groups”. In: *Thin groups and superstrong approximation*. Vol. 61. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, pp. 343–362. MR: 3220897 (cit. on p. 1050).

Carlos T. Simpson (1991). “Nonabelian Hodge theory”. In: *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*. Math. Soc. Japan, Tokyo, pp. 747–756. MR: 1159261 (cit. on p. 1032).

– (1992). “Higgs bundles and local systems”. *Inst. Hautes Études Sci. Publ. Math.* 75, pp. 5–95 (cit. on p. 1032).

Tobias Strubel (2015). "Fenchel-Nielsen coordinates for maximal representations". *Geom. Dedicata* 176, pp. 45–86. MR: 3347573 (cit. on p. 1045).

Zhe Sun (2017a). "Rank n swapping algebra for the $\mathrm{PSL}(n,\mathbb{R})$ Hitchin component". *Int. Math. Res. Not. IMRN* 2, pp. 583–613. MR: 3658146 (cit. on p. 1032).

– (2017b). "Rank n swapping algebra for PGL_n Fock-Goncharov X moduli space". arXiv: 1503.00918 (cit. on p. 1032).

Zhe Sun, Anna Wienhard, and Tengren Zhang (2017). "Flows on the PSL(V)-Hitchin component". arXiv: 1709.03580 (cit. on pp. 1046, 1047).

Zhe Sun and Tengren Zhang (2017). "The Goldman symplectic form on the PSL(V)-Hitchin component". arXiv: 1709.03589 (cit. on pp. 1046, 1047).

Nicolas Tholozan (2017). "Volume entropy of Hilbert metrics and length spectrum of Hitchin representations into $\mathrm{PSL}(3,\mathbb{R})$". *Duke Math. J.* 166.7, pp. 1377–1403. MR: 3649358 (cit. on p. 1048).

Domingo Toledo (1987). "Nonexistence of certain closed complex geodesics in the moduli space of curves". *Pacific J. Math.* 129.1, pp. 187–192. MR: 901262 (cit. on p. 1051).

– (1989). "Representations of surface groups in complex hyperbolic space". *J. Differential Geom.* 29.1, pp. 125–133 (cit. on p. 1036).

È. B. Vinberg and V. G. Kac (1967). "Quasi-homogeneous cones". *Mat. Zametki* 1, pp. 347–354. MR: 0208470 (cit. on p. 1050).

Nicolas Vlamis and Andrew Yarmola (2017). "Basmajian's identity in higher Teichmüller-Thurston theory". *Journal of Topology.* 10.3, pp. 744–764 (cit. on p. 1049).

Anna Wienhard (2006). "The action of the mapping class group on maximal representations". *Geom. Dedicata* 120, pp. 179–191 (cit. on p. 1041).

Anna Wienhard and Tengren Zhang (2018). "Deforming convex real projective structures". *Geom. Dedicata* 192, pp. 327–360. MR: 3749433 (cit. on p. 1047).

Scott Wolpert (1983). "On the symplectic geometry of deformations of a hyperbolic surface". *Ann. of Math. (2)* 117.2, pp. 207–234. MR: 690844 (cit. on p. 1046).

Tengren Zhang (2015a). "Degeneration of Hitchin representations along internal sequences". *Geom. Funct. Anal.* 25.5, pp. 1588–1645. MR: 3426063 (cit. on pp. 1046, 1048).

– (2015b). "The degeneration of convex $\mathbb{RP}^2$ structures on surfaces". *Proc. Lond. Math. Soc. (3)* 111.5, pp. 967–1012. MR: 3477227 (cit. on p. 1048).

Received 2018-03-01.

Anna Wienhard

Ruprecht-Karls Universität Heidelberg
Mathematisches Institut
Im Neuenheimer Feld 205
69120 Heidelberg
Germany

and

HITS gGmbH
Heidelberg Institute for Theoretical Studies
Schloss-Wolfsbrunnenweg 35
69118 Heidelberg
Germany
wienhard@mathi.uni-heidelberg.de

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1059–1080)

K-THEORY AND ACTIONS ON EUCLIDEAN RETRACTS

Arthur Bartels

Abstract

This note surveys axiomatic results for the Farrell-Jones Conjecture in terms of actions on Euclidean retracts and applications of these to $\mathrm{GL}_n(\mathbb{Z})$, relative hyperbolic groups and mapping class groups.

Introduction

Motivated by surgery theory Hsiang [1984] made a number of influential conjectures about the K-theory of integral group rings $\mathbb{Z}[G]$ for torsion free groups G. These conjectures often have direct implications for the classification theory of manifolds of dimension ≥ 5. A good example is the following. An h-cobordism is a compact manifold W that has two boundary components M_0 and M_1 such that both inclusions $M_i \to W$ are homotopy equivalences. The Whitehead group $\mathrm{Wh}(G)$ is the quotient of $K_1(\mathbb{Z}[G])$ by the subgroup generated by the canonical units $\pm g$, $g \in G$. Associated to an h-cobordism is an invariant, the Whitehead torsion, in $\mathrm{Wh}(G)$, where G is the fundamental group of W. A consequence of the s-cobordism theorem is that for $\dim W \geq 6$, an h-cobordism W is trivial (i.e., isomorphic to a product $M_0 \times [0,1]$) iff its Whitehead torsion vanishes. Hsiang conjectured that for G torsion free $\mathrm{Wh}(G) = 0$, and thus that in many cases h-cobordisms are products.

The Borel conjecture asserts that closed aspherical manifolds are topologically rigid, i.e., that any homotopy equivalence to another closed manifold is homotopic to a homeomorphism. The last step in proofs of instances of this conjecture via surgery theory uses a vanishing result for $\mathrm{Wh}(G)$ to conclude that an h-cobordism is a product and that therefore the two boundary components are homeomorphic.

Farrell and Jones [1986] pioneered a method of using the geodesic flow on non-positively curved manifolds to study these conjectures. This created a beautiful connection between

The work has been supported by the SFB 878 in Münster.
MSC2010: primary 18F25; secondary 19G24, 20F67, 20F65.
Keywords: Farrell-Jones Conjecture, K- and L-theory of group rings.

K-theory and dynamics that led Farrell and Jones [1993b], among many other results, to a proof of the Borel Conjecture for closed Riemannian manifolds of non-positive curvature of dimension ≥ 5. Moreover, Farrell and Jones [1993a] formulated (and proved many instances of) a conjecture about the structure of the algebraic K-theory (and L-theory) of group rings, even in the presence of torsion in the group. Roughly, the Farrell-Jones Conjecture states that the main building blocks for the K-theory of $\mathbb{Z}[G]$ is the K-theory of $\mathbb{Z}[V]$ where V varies over the family of virtually cyclic subgroups of G. It implies a number of other conjectures, among them Hsiang's conjectures, the Borel Conjecture in dimension ≥ 5, the Novikov Conjecture on the homotopy invariant of higher signatures, Kaplansky's conjecture about idempotents in group rings, see Lück [2010] for a summary of these and other applications.

My goal in this note is twofold. The first goal is to explain a condition formulated in terms of existence of certain actions of G on Euclidean retracts that implies the Farrell-Jones Conjecture for G. This condition was developed in A. Bartels and Lück [2012c] and A. Bartels, Lück, and Reich [2008b] where the connection between K-theory and dynamics has been extended beyond the context of Riemannian manifolds to prove the Farrell-Jones Conjecture for hyperbolic and CAT(0)-groups. The second goal is to outline how this condition has been used in joint work with Lück, Reich and Rüping and with Bestvina to prove the Farrell-Jones Conjecture for $\mathrm{GL}_n(\mathbb{Z})$ and mapping class groups. A common difficulty for both families of groups is that their natural proper actions (on the associated symmetric space, respectively on Teichmüller space) is not cocompact. In both cases the solution depends on a good understanding of the action away from cocompact subsets and an induction on a complexity of the groups. As a preparation for mapping class groups we also discuss relatively hyperbolic groups.

The Farrell-Jones Conjecture has a prominent relative, the Baum-Connes Conjecture for topological K-theory of group C^*-algebras Baum and Connes [2000] and Baum, Connes, and Higson [1994]. The two conjectures are formally very similar, but methods of proofs are different. In particular, the conditions discussed in Section 2 are not known to imply the Baum-Conjecture. The classes of groups for which the two conjectures are known differ. For example, by work of Kammeyer, Lück, and Rüping [2016] all lattice in Lie groups satisfy the Farrell-Jones Conjecture; despite Lafforgue [2002] positive results for many property T groups, the Baum-Connes Conjecture is still a challenge for $\mathrm{SL}_3(\mathbb{Z})$. Wegner [2015] proved the Farrell-Jones Conjecture for all solvable groups, but the case of amenable (or just elementary amenable) groups is open; in contrast Higson and Kasparov [2001] proved the Baum-Connes Conjecture for all a-T-menable groups, a class of groups that contains all amenable groups. On the other hand, hyperbolic groups satisfy both conjectures. See Mineyev and Yu [2002] and Lafforgue [2012] for the Baum-Connes Conjecture and, as mentioned above, A. Bartels and Lück [2012c] and A. Bartels, Lück, and Reich [2008b] for the Farrell-Jones Conjecture. For a more comprehensive summary of the current status of the Farrell-Jones Conjecture the reader is directed to Lück [2017] and Reich and Varisco [2017].

Acknowledgments. It is a pleasure to thank my teachers, coauthors, and students for the many things they taught me.

1 The Formulation of the Farrell-Jones Conjecture

Classifying spaces for families. A family $\mathcal{F}$ of subgroups of a group G is a non-empty collection of subgroups that is closed under conjugation and subgroups. Examples are the family Fin of finite subgroups and the family VCyc of virtually cyclic subgroups (i.e., of subgroups containing a cyclic subgroup as a subgroup of finite index). For any family of subgroups of G there exists a G-CW-complex $E_{\mathcal{F}}G$ with the following property: if E is any other G-CW-complex such that all isotropy groups of E belong to $\mathcal{F}$, then there is a up to G-homotopy unique G-map $E \to E_{\mathcal{F}}G$. This space is not unique, but it is unique up to G-homotopy equivalence. Informally one may think about $E_{\mathcal{F}}G$ as a space that encodes the group G relative to all subgroups from $\mathcal{F}$. Often there are interesting geometric models for this space, in particular for $\mathcal{F} = \text{Fin}$. More information about this space can be found for example in Lück [2005]. An easy way to construct $E_{\mathcal{F}}G$ is as the infinite join $*_{i=0}^{\infty}(\coprod_{F\in\mathcal{F}} G/F)$. If $\mathcal{F}$ is closed under supergroups of finite index (i.e., if $F \in \mathcal{F}$ is a subgroup of finite index in F', then also $F' \in \mathcal{F}$), then the full simplicial complex on $\coprod_{F\in\mathcal{F}} G/F$ is also a model for $E_{\mathcal{F}}G$; we will denote this model later by $\Delta_{\mathcal{F}}(G)$.

The formulation of the conjecture. The original formulation of the Farrell and Jones [1993a] used homology with coefficients in stratified and twisted Ω-spectra. Here we use the equivalent Hambleton and Pedersen [2004] formulation developed by Davis and Lück [1998]. Given a ring R, Davis-Lück construct a homology theory $X \mapsto H_*^G(X;\mathbf{K}_R)$ for G-spaces with the property that $H_*^G(G/H;\mathbf{K}_R) \cong K_*(R[H])$.

Let $\mathcal{F}$ be a family of subgroups of the group G. Consider the projection map $E_{\mathcal{F}}G \to G/G$ to the one-point G-space G/G. It induces the $\mathcal{F}$-assembly map

$$\alpha_{\mathcal{F}}^G \colon H_*^G(E_{\mathcal{F}};\mathbf{K}_R) \to H_*^G(G/G;\mathbf{K}_R) \cong K_*(R[G]).$$

Conjecture 1.1 (Farrell-Jones Conjecture)**.** *For any group G and any ring R the assembly map* α_{VCyc}^G *is an isomorphism.*

This version of the conjecture has been stated in A. Bartels, Farrell, Jones, and Reich [2004]. The original formulation of Farrell and Jones [1993a] considered only the integral group ring $\mathbb{Z}[G]$. Moreover, Farrell and Jones wrote that they regard this and related conjectures *only as estimates which best fit the know data at this time*. However, the conjecture is still open and does still fit with all known data today.

Transitivity principle. Informally one can view the statement that the assembly map $\alpha^G_{\mathcal{F}}$ is an isomorphism for a group G and a ring R as the statement that $K_*(R[G])$ can be assembled from $K_*(R[F])$ for all $F \in \mathcal{F}$ (and group homology). If $\mathcal{V}$ is a family of subgroups of G that contains all subgroups from $\mathcal{F}$, then one can apply this slogan in two steps, for G relative to $\mathcal{V}$ and for each $V \in \mathcal{V}$ relative to the $F \in \mathcal{F}$ with $F \subseteq V$. The implementation of this is the following transitivity principle.

Theorem 1.2 (Farrell and Jones [1993a] and Lück and Reich [2005])**.** *For $V \in \mathcal{V}$ set $\mathcal{F}_V := \{F \mid F \in \mathcal{F}, F \subseteq V\}$. Assume that for all $V \in \mathcal{V}$ the assembly map $\alpha^V_{\mathcal{F}_V}$ is an isomorphism. Then $\alpha^G_{\mathcal{F}}$ is an isomorphism iff $\alpha^G_{\mathcal{V}}$ is an isomorphism.*

Twisted coefficients. Often it is beneficial to study more flexible generalizations of Conjecture 1.1. Such a generalization is the Fibred Isomorphism Conjecture of Farrell and Jones [1993a]. An alternative is the Farrell-Jones Conjecture with coefficients in additive categories A. Bartels and Reich [2007], here one allows additive categories with an action of a group G instead of just a ring as coefficients. This version of the conjecture applies in particular to twisted group rings. These generalizations of the Conjecture have better inheritance properties. Two of these inheritance property are stability under directed colimits of groups, and stability under taking subgroups. For a summary of the inheritance properties see Reich and Varisco [2017, Thm. 27(2)]. Often proofs of cases of the Farrell-Jones Conjecture use these inheritance properties in inductions or to reduce to special cases. We will mean by the statement that G satisfies the Farrell-Jones Conjecture relative to $\mathcal{F}$, that the assembly map $\alpha^G_{\mathcal{F}}$ is bijective for all additive categories $\mathcal{A}$ with G-action. However, this as a technical point that can be safely ignored for the purpose of this note.

Other theories. The Farrell-Jones Conjecture for K-theory discussed so far has an analog in L-theory, as it appears in surgery theory. For some of the applications mentioned before this is crucial. For example, the Borel Conjecture for a closed aspherical manifold M of dimensions ≥ 5 holds if the fundamental group of M satisfies both the K-and L-theoretic Farrell-Jones Conjecture. However, proofs of the Farrell-Jones Conjecture in K- and L-theory are by now very parallel. Recently, the techniques for the Farrell-Jones Conjecture in K- and L-theory have been extended to also cover Waldhausen's A-theory Enkelmann, Lück, Pieper, Ullmann, and Winges [2016], Kasprowski, Ullmann, Wegner, and Winges [2018], and Ullmann and Winges [2015]. In particular, the conditions we will discuss in Section 2 are now known to imply the Farrell-Jones Conjecture in all three theories.

2 Actions on compact spaces

Amenable actions and exact groups.

Definition 2.1 (Almost invariant maps)**.** Let X, E be G-spaces where E is equipped with a G-invariant metric d. We will say that a sequence of maps $f_n\colon X \to E$ is almost G-equivariant if for any $g \in G$

$$\sup_{x \in X} d(f_n(gx), gf_n(x)) \to 0 \quad \text{as} \quad n \to \infty.$$

For a discrete group G we equip the space $\operatorname{Prob}(G)$ of probability measures on G with the metric it inherits as subspace of $l^1(G)$. This metric generates the topology of pointwise convergence on $\operatorname{Prob}(G)$. We recall the following definition.

Definition 2.2. An action of a group G on a compact space X is said to be amenable if there exists a sequence of almost equivariant maps $X \to \operatorname{Prob}(G)$.

A group is amenable iff its action on the one point space is amenable. Groups that admit an amenable action on a compact Hausdorff space are said to be *exact* or *boundary amenable*. The class of exact groups contains all amenable groups, hyperbolic groups Adams [1994], and all linear groups Guentner, Higson, and Weinberger [2005]. Other prominent groups that are known to be exact are mapping class groups Hamenstädt [2009] and Kida [2008] and the group of outer automorphisms of free groups Bestvina, Guirardel, and Horbez [2017]. The Baum-Connes assembly map is split injective for all exact groups Higson [2000] and Yu [2000]. This implies the Novikov conjecture for exact groups. This is an analytic result for the Novikov conjecture, in the sense that it has no known proof that avoids the Baum-Connes Conjecture. There is no corresponding injectivity result for assembly maps in algebraic K-theory. For a survey about amenable actions and exact groups see Ozawa [2006].

Finite asymptotic dimension. Results for assembly maps in algebraic K-theory and L-theory often depend on a finite dimensional setting; the space of probability measures has to be replaced with a finite dimensional space. We write $\Delta(G)$ for the full simplicial complex with vertex set G and $\Delta^{(N)}(G)$ for its N-skeleton. The space $\Delta(G)$ can be viewed as the space of probability measures on G with finite support. We equip $\Delta(G)$ with the l^1-metric; this is the metric it inherits from $\operatorname{Prob}(G)$.

Definition 2.3 (N-amenable action)**.** We will say that an action of a group G on a compact space X is N-amenable if there exists a sequence of almost equivariant maps $X \to \Delta^{(N)}(G)$.

The natural action of a countable group G on its Stone-Čech compactification βG is N-amenable iff the asymptotic dimension of G is at most N Guentner, Willett, and Yu [2017, Thm. 6.5]. This condition (for any N) also implies exactness and therefore the Novikov conjecture Higson and Roe [2000]. For groups G of finite asymptotic dimension for which in addition the classifying space BG can be realized as a finite CW-complex, there is an alternative argument for the Novikov Conjecture Yu [1998] that has been translated to integral injectivity results for assembly maps in algebraic K-theory and L-theory A. C. Bartels [2003] and Carlsson and Goldfarb [2004]. These injectivity results have seen far reaching generalizations to groups of finite decomposition complexity Guentner, Tessera, and Yu [2012], Kasprowski [2015], and Ramras, Tessera, and Yu [2014].

***N*-$\mathfrak{F}$-amenable actions.** Constructions of transfer maps in algebraic K-theory and L-theory often depend on actions on spaces that are much nicer than βG. A good class of spaces to use for the Farrell-Jones Conjecture are Euclidean retracts, i.e., compact spaces that can be embedded as a retract in some $\mathbb{R}^n$. Brouwer's fixed point theorem implies that for an action of a group on an Euclidean retract any cyclic subgroup will have a fixed point. It is not difficult to check that this obstructs the existence of almost equivariant maps to $\Delta^{(N)}$ (assuming G contains an element of infinite order). Let $\mathfrak{F}$ be a family of subgroups of G that is closed under taking supergroups of finite index. Let $S := \coprod_{F \in \mathfrak{F}} G/F$ be the set of all left cosets to members of $\mathfrak{F}$. Let $\Delta_{\mathfrak{F}}(G)$ be the full simplicial complex on S and $\Delta_{\mathfrak{F}}^{(N)}(G)$ be its N-skeleton. We equip $\Delta_{\mathfrak{F}}(G)$ with the l^1-metric.

Definition 2.4 (N-$\mathfrak{F}$-amenable action)**.** We will say that an action of G on a compact space X is N-$\mathfrak{F}$-amenable if there exists a sequence of almost equivariant maps $X \to \Delta_{\mathfrak{F}}^{(N)}(G)$. If an action is N-$\mathfrak{F}$-amenable for some $N \in \mathbb{N}$, then we say that it is finitely $\mathfrak{F}$-amenable.

Remark 2.5. Let X be a G-CW-complex with isotropy groups in $\mathfrak{F}$ and of dimension $\leq N$. As $\Delta_{\mathfrak{F}}(G)$ is a model for $E_{\mathfrak{F}}G$ we obtain a cellular G-map $f : X \to \Delta_{\mathfrak{F}}^{(N)}(G)$; this map is also continuous for the l^1-metric. In particular, the constant sequence $f_n \equiv f$ is almost equivariant. Therefore one can view N-$\mathfrak{F}$-amenability for G-spaces as a relaxation of the property of being a G-CW-complex with isotropy in $\mathfrak{F}$ and of dimension $\leq N$.

This relaxation is necessary to obtain compact examples and reasonably small $\mathfrak{F}$: If a G-CW-complex is compact, then it has only finitely many cells. In particular, for each cell the isotropy group has finite index in G, so $\mathfrak{F}$ would have to contain subgroups of finite index in G.

Theorem 2.6 (A. Bartels and Lück [2012c] and A. Bartels, Lück, and Reich [2008b])**.** *Suppose that G admits a finitely $\mathfrak{F}$-amenable action on a Euclidean retract. Then G satisfies the Farrell-Jones Conjecture relative to $\mathfrak{F}$.*

Remark 2.7. The proof of Theorem 2.6 depends on methods from controlled topology/algebra that have a long history. An introduction to controlled algebra is given in Pedersen [2000]; an introduction to the proof of Theorem 2.6 can be found in A. Bartels [2016]. Here we only sketch a very special case, where these methods are not needed.

Assume that the Euclidean retract is a G-CW-complex X. As pointed out in Remark 2.5 this forces $\mathfrak{F}$ to contain subgroups of finite index in G. As X is contractible, the cellular chain complex of X provides a finite resolution C_* over $\mathbb{Z}[G]$ of the trivial G-module $\mathbb{Z}$. Note that in each degree $C_k = \bigoplus \mathbb{Z}[G/F_i]$ is a finite sum of permutation modules with $F_i \in \mathfrak{F}$ and F_i of finite index in G. For a finitely generated projective $R[G]$-module P we obtain a finite resolution $C_* \otimes_{\mathbb{Z}} P$ of P. Each module in the resolution is a finite sum of modules of the form $\mathbb{Z}[G/F] \otimes_{\mathbb{Z}} P$ with $F \in \mathfrak{F}$ and of finite index in G. Here $\mathbb{Z}[G/F] \otimes_{\mathbb{Z}} P$ is equipped with the diagonal G-action and can be identified with the $R[G]$-module obtained by first restricting P to an $R[F]$-module and then inducing back up from $R[F]$ to $R[G]$. In particular, $\big[\mathbb{Z}[G/F] \otimes_{\mathbb{Z}} P\big] \in K_0(R[G])$ is in the image of the assembly map relative to the family $\mathfrak{F}$. It follows that $[P] = \sum_k (-1)^k [C_k \otimes_{\mathbb{Z}} P]$ is also in the image. Therefore the assembly map $H_0(E_{\mathfrak{F}} G; \mathbf{K}_R) \to K_0(R[G])$ is surjective. (This argument did not use that $\mathfrak{F}$ is closed under supergroups of finite index.)

Example 2.8. Let G be a hyperbolic group. Its Rips complex can be compactified to a Euclidean retract Bestvina and Mess [1991]. The natural action of G on this compactification is finitely VCyc-amenable A. Bartels, Lück, and Reich [2008a].

To obtain further examples of finitely $\mathfrak{F}$-amenable actions on Euclidean retracts, it is helpful to replace VCyc with a larger family of subgroups $\mathfrak{F}$. Groups that act acylindrically hyperbolic on a tree admit finitely $\mathfrak{F}$-amenable actions on Euclidean retracts where $\mathfrak{F}$ is the family of subgroups that is generated by the virtually cyclic subgroups and the isotropy groups for the original action on the tree Knopf [2017]. Relative hyperbolic groups and mapping class groups are discussed in Section 4.

Remark 2.9. A natural question is which groups admit finitely VCyc-amenable actions on Euclidean retracts. A necessary condition for an action to be finitely VCyc-amenable is that all isotropy groups of the action are virtually cyclic. Therefore, a related question is which groups admit actions on Euclidean retracts such that all isotropy groups are virtually cyclic. The only groups admitting such actions that I am aware of are hyperbolic groups. In fact, I do not even know whether or not the group $\mathbb{Z}^2$ admits an action on a Euclidean retract (or on a disk) such that all isotropy groups are virtually cyclic. There are actions of $\mathbb{Z}^2$ on disks without a global fixed point. This is a consequence of Oliver's analysis of actions of finite groups on disks Oliver [1975]. On the other hand, there are finitely generated groups for which all actions on Euclidean retracts have a global fixed point Arzhantseva, Bridson, Januszkiewicz, Leary, Minasyan, and Świ tkowski [2009].

Homotopy actions. There is a generalization of Theorem 2.6 using homotopy actions. In order to be applicable to higher K-theory these actions need to be homotopy coherent. The passage from strict actions to homotopy actions is already visible in the work of Farrell-Jones where it corresponds to the passage from the asymptotic transfer used for negatively curved manifolds Farrell and Jones [1986] to the focal transfer used for non-positively curved manifolds Farrell and Jones [1993b].

Definition 2.10 (Vogt [1973] and Wegner [2012])**.** A *homotopy coherent action* of a group G on a space X is a continuous map

$$\Gamma\colon \coprod_{j=0}^{\infty}((G\times[0,1])^j\times G\times X) \to X$$

such that

$$\Gamma(g_k,t_k,\dots,t_1,g_0,x) = \begin{cases} \Gamma(g_k,\dots,g_j,\Gamma(g_{j-1},\dots,x)) & t_j=0 \\ \Gamma(g_k,\dots,g_jg_{j-1},\dots,x) & t_j=1 \\ \Gamma(g_k,\dots,t_2,g_1,x) & g_0=e, 0<k \\ \Gamma(g_k,\dots,t_{j+1}t_j,\dots,g_0,x) & g_j=e, 1\le j<k \\ \Gamma(g_{k-1},\dots,t_1,g_0,x) & g_k=e, 0<k \\ x & g_0=e, k=0 \end{cases}$$

Here $\Gamma(g,-)\colon X \to X$ should be thought of the action of g on X, the map $\Gamma(g,-,h,-)\colon [0,1]\times X \to X$ is a homotopy from $\Gamma(g,-)\circ\Gamma(h,-)$ to $\Gamma(gh,-)$ and the remaining data in Γ encodes higher coherences.

In order to obtain sequences of almost equivariant maps for homotopy actions it is useful to also allow the homotopy action to vary.

Definition 2.11 (N-$\mathfrak{F}$-amenability for homotopy coherent actions)**.** A sequence of homotopy coherent actions (Γ_n, X_n) of a group G is said to be N-$\mathfrak{F}$-amenable if there exists a sequence of continuous maps $f_n\colon X_n \to \Delta^{(N)}_{\mathfrak{F}}(G)$ such that for all k and all $g_k,\dots,g_0 \in G$

$$\sup_{x\in X, t_k,\dots,t_1\in[0,1]} d(f_n(\Gamma(g_k,t_k,\dots,t_1,g_0,x), g_k\cdots g_0 f_n(x)) \to 0 \quad \text{as} \quad n\to\infty.$$

Theorem 2.12 (A. Bartels and Lück [2012c] and Wegner [2012])**.** *Suppose that G admits a sequence of homotopy coherent actions on Euclidean retracts of uniformly bounded dimension that is finitely $\mathfrak{F}$-amenable. Then G satisfies the Farrell-Jones Conjecture relative to $\mathfrak{F}$.*

Remark 2.13. Groups satisfying the assumptions of Theorem 2.12 are said to be homotopy transfer reducible in Enkelmann, Lück, Pieper, Ullmann, and Winges [2016]. The original formulations of Theorems 2.6 and 2.12 were not in terms of almost equivariant maps, but in terms of certain open covers of $G \times X$.

We recall here the formulation used for actions. (The formulation for homotopy actions is more cumbersome.) A subset U of a G-space is said to be an $\mathfrak{F}$-subset if there is $F \in \mathfrak{F}$ such that $gU = U$ for all $g \in F$ and $U \cap gU = \varnothing$ for all $g \in G \setminus F$. A collection $\mathfrak{U}$ of subsets is said to be G-invariant if $gU \in \mathfrak{U}$ for all $g \in G$, $U \in \mathfrak{U}$. If no point is contained in more than $N+1$ members of $\mathfrak{U}$, then $\mathfrak{U}$ is said to be of order (or dimension) $\leq N$. A G-invariant cover by open $\mathfrak{F}$-subsets is said to be an $\mathfrak{F}$-cover.

For a compact G-space X we equip now $G \times X$ with the diagonal G-action. For $S \subseteq G$ finite an $\mathfrak{F}$-cover of $G \times X$ is said to be S-wide (in the G-direction) if

$$\forall\, (g, x) \in G \times X \;\; \exists\, U \in \mathfrak{U} \;\text{ such that }\; gS \times \{x\} \subseteq U.$$

Then the action of G on X is N-$\mathfrak{F}$-amenable iff for any $S \subset G$ finite there exists an S-wide $\mathfrak{F}$-cover $\mathfrak{U}$ for $G \times X$ of dimension at most N Guentner, Willett, and Yu [2017, Prop. 4.5]. A translation from covers to maps is also used in A. Bartels and Lück [2012c], A. Bartels, Lück, and Reich [2008b], and Wegner [2012]. From the point of view of covers (and because of the connection to the asymptotic dimension) it is natural to think of the N in N-$\mathfrak{F}$-amenable as a kind of dimension for the action of G on X, see Guentner, Willett, and Yu [2017] and Sawicki [2017].

A further difference between the formulations used above and in the references given is that the conditions on the topology of X are formulated differently, but certainly Euclidean retracts satisfy the condition from A. Bartels and Lück [2012c].

Example 2.14. Theorem 2.12 applies to CAT(0)-groups where $\mathfrak{F} = \mathrm{VCyc}$ is the family of virtually cyclic subgroups A. Bartels and Lück [2012b] and Wegner [2012]. An application of Theorem 2.12 to $\mathrm{GL}_n(\mathbb{Z})$ will be discussed in Section 4.

Remark 2.15 (The Farrell-Hsiang method). There are interesting groups for which one can deduce the Farrell-Jones Conjecture using Theorems 2.6 (or 2.12) and inheritance properties. However, it is not clear that these methods can account for all that is currently known. A third method, going back to work of Farrell and Hsiang [1978], combines induction results for finite groups Dress [1975] and Swan [1960] with controlled topology/algebra. An axiomatization of this method is given in A. Bartels and Lück [2012a]. In important part of the proof of the Farrell-Jones Conjecture for solvable groups Wegner [2015] is a combination of this method with Theorem 2.12.

Remark 2.16 (Trace methods). The K-theory Novikov Conjecture concerns injectivity of assembly maps in algebraic K-theory, i.e., lower bounds for the algebraic K-theory of group rings. For the integral group ring of groups, that are only required to satisfy a mild homological finiteness assumption, trace methods have been used by Bökstedt, Hsiang, and Madsen [1993] and Lück, Reich, Rognes, and Varisco [2017] to obtain rational injectivity results. The latter result in particular yields interesting lower bounds for Whitehead groups. For the group ring over the ring of Schatten class operators Yu [2017] proved rational injectivity of the Farrell-Jones assembly map for all groups. This is the only result I am aware of for the Farrell-Jones Conjecture that applies to all groups!

3 Flow spaces

The construction of almost equivariant maps often uses the dynamic of a flow associated to the situation.

Definition 3.1. A *flow space* for a group G is a metric space FS equipped with a flow Φ and an isometric G-action where the flow and the G-action commute. For $\alpha > 0$, $\delta > 0$, $c, c' \in FS$ we write

$$d_{fol}(c, c') < (\alpha, \delta)$$

to mean that there is $t \in [-\alpha, \alpha]$ such that $d(\Phi_t(c), c') < \delta$.

Example 3.2. Let G be the fundamental group of a Riemannian manifold M. Then the sphere bundle $S\tilde{M}$ equipped with the geodesic flow is a flow space for the fundamental group of M. For manifolds of negative or non-positive curvature this flow space is at the heart of the connection between K-theory and dynamics used to great effect by Farrell-Jones.

This example has generalizations to hyperbolic groups and CAT(0)-groups. For hyperbolic groups Mineyev's symmetric join is a flow space Mineyev [2005]. Alternatively, it is possible to use a coarse flow space for hyperbolic groups, see Remark 3.7 below. For groups acting on a CAT(0)-space a flow space has been constructed in A. Bartels and Lück [2012b]. It consists of all parametrized geodesics in the CAT(0)-space (technically all generalized geodesics) and the flow acts by shifting the parametrization.

Almost equivariant maps often arise as compositions

$$X \xrightarrow{\varphi} FS \xrightarrow{\psi} \Delta^{(N)}_{\mathfrak{F}}(G),$$

where the first map is almost equivariant in an (α, δ)-sense, and the second map is G-equivariant and contracts (α, δ)-distances to ε-distances. The following Lemma summarizes this strategy.

Lemma 3.3. *Let X be a G-space, where G is a countable group. Let $N \in \mathbb{N}$. Assume that there exists a flow space FS satisfying the following two conditions.*

(A) *For any finite subset S of G there is $\alpha > 0$ such that for any $\delta > 0$ there is a continuous map $\varphi\colon X \to FS$ such that for $x \in X$, $g \in S$ we have*

$$d_{fol}(\varphi(gx), g\varphi(x)) < (\alpha, \delta).$$

(B) *For any $\alpha > 0$, $\varepsilon > 0$ there are $\delta > 0$ and a continuous G-map $\psi : FS \to \Delta^{(N)}_{\mathfrak{F}}(G)$ such that*

$$d_{fol}(c, c') < (\alpha, \delta) \quad \Longrightarrow \quad d(\psi(c), \psi(c')) < \varepsilon$$

holds for all $c, c' \in FS$.

Then the action of G on X is N-$\mathfrak{F}$-amenable.

Proof. Let $S \subset G$ be finite and $\varepsilon > 0$. We need to construct a map $f : X \to \Delta^{(N)}_{\mathfrak{F}}(G)$ for which $d(f(gx), gf(x)) < \varepsilon$ for all $x \in X$, $g \in S$. Let α be as in (A) with respect to S. Choose now $\delta > 0$ and a G-map $\psi : FS \to \Delta^{(N)}_{\mathfrak{F}}(G)$ as in (B). Next choose $\varphi\colon X \to FS$ as in (A) with respect to this $\delta > 0$. Then $f := \psi \circ \varphi$ has the required property. □

Remark 3.4 (On the constructions of φ). Maps $\varphi\colon X \to FS$ as in condition (A) in Lemma 3.3 can in negatively or non-positively curved situations often be constructed using dynamic properties of the flow. We briefly illustrate this in a case already considered by Farrell and Jones.

Let G be the fundamental group of a closed Riemannian manifold of strict negative sectional curvature M. Let $\tilde{M}$ be its universal cover and S_∞ the sphere at infinity for $\tilde{M}$. The action of G on $\tilde{M}$ extends to S_∞. For each $x \in \tilde{M}$ there is a canonical identification between the unit tangent vectors at x and S_∞: every unit tangent vector v at x determines a geodesic ray c starting in x, the corresponding point $\xi \in S_\infty$ is $c(\infty)$. One say that v points to ξ. The geodesic flow Φ_t on $S\tilde{M}$ has the following property. Suppose that v and v' are unit tangent vectors at x and x' pointing to the same point in S_∞. Then $d_{fol}(\Phi_t(v), \Phi_t(v')) < (\alpha, \delta_t)$ where α depends only on $d(x, x_0)$ and $\delta_t \to 0$ uniformly in v, v' (still depending on $d(x, x_0)$). This statement uses strict negative curvature. (For closed manifolds of non-positive sectional curvature the vector v' has to be chosen more carefully depending on v and t; this necessitates the use of the focal transfer Farrell and Jones [1993b] respectively the use of homotopy coherent actions.)

This contracting property of the geodesic flow can be translated into the construction of maps as in (A). Fix a point $x_0 \in \tilde{M}$. Define $\varphi_0\colon S_\infty \to S\tilde{M}$ by sending ξ to the unit tangent vector at x_0 pointing to ξ. For $t \geq 0$ define $\varphi_t(\xi) := \Phi_t(\varphi_0(\xi))$. Using the contracting property of the geodesic flow is not difficult to check that for any $g \in G$ there is $\alpha > 0$ (roughly $\alpha = d(gx_0, x_0)$) such that for any $\delta > 0$ there is t_0 satisfying

$$\forall t \geq t_0,\ \forall \xi \in S_\infty \ \ d_{fol}(\varphi_t(g\xi), g\varphi_t(\xi)) < (\alpha, \delta).$$

Remark 3.5. Of course the space S_∞ used in Remark 3.4 is not contractible and therefore not a Euclidean retract. But the compactification $\tilde{M} \cup S_\infty$ of $\tilde{M}$ is a disk, in particular a Euclidean retract. As $\tilde{M}$ has the homotopy type of a free G-CW-complex, there is even a G-equivariant map $\tilde{M} \to \Delta^{(d)}(G)$ where d is the dimension of M. In particular, the action of G on $\tilde{M}$ is d-amenable. It is not difficult to combine the two statements to deduce that the action of G on $\tilde{M} \cup S_\infty$ is finitely $\mathfrak{F}$-amenable. This is best done via the translation to open covers of $G \times (\tilde{M} \cup S_\infty)$ discussed in Remark 2.13, see for example Sawicki [2017].

An important point in the formulation of condition (B) is the presence of $\delta > 0$ uniform over *FS*. If the action of G on *FS* is cocompact, then a version of the Lebesgue Lemma guarantees the existence of some uniform $\delta > 0$, i.e., it suffices to construct $\psi : FS \to \Delta_{\mathfrak{F}}^{(N)}(G)$ such that $d(\psi(c), \psi(\Phi_t(c)) < \varepsilon$ for all $t \in [-\alpha, \alpha]$, $c \in FS$.

Remark 3.6 (Long thin covers of *FS*). Maps φ as in condition (B) of Lemma 3.3 are best constructed as maps associated to long thin covers of the flow space. These long thin covers are an alternative to the long thin cell structures employed by Farrell and Jones [1986].

An open cover $\mathfrak{U}$ of *FS* is said to be an α-long cover for *FS* if for each $c \in FS$ there is $U \in \mathfrak{U}$ such that

$$\Phi_{[-\alpha,\alpha]}(c) \subseteq U.$$

It is said to be α-long and δ-thick if for each $c \in FS$ there is $U \in \mathfrak{U}$ containing the δ-neighborhood of $\Phi_{[-\alpha,\alpha]}(c)$. The construction of maps $FS \to \Delta_{\mathfrak{F}}^{(N)}(G)$ as in condition (B) of Lemma 3.3 amounts to finding for given α an $\mathfrak{F}$-cover $\mathfrak{U}$ of *FS* of dimension at most N that is α-long and δ-thick for some $\delta > 0$ (depending on α). For cocompact flow spaces such covers can be constructed in relatively great generality A. Bartels, Lück, and Reich [2008a] and Kasprowski and Rüping [2017]. Cocompactness is used to guarantee δ-thickness. For not cocompact flow spaces on can still find α-long covers, but without a uniform thickness, they do not provide the maps needed in (B).

Remark 3.7 (Coarse flow space). We outline the construction of the coarse flow space from A. Bartels [2017] for a hyperbolic group G. Let Γ be a Cayley graph for G. The vertex set of Γ is G. Adding the Gromov boundary to G we obtain the compact space $\overline{G} = G \cup \partial G$. Assume that Γ is δ-hyperbolic. The *coarse flow space CF* consists of all triples (ξ_-, v, ξ_+) with $\xi_\pm \in \overline{G}$ and $v \in G$ such that there is some geodesic from ξ_- to ξ_+ in Γ that passes v within distance $\leq \delta$. Informally, v coarsely belongs to a geodesic from ξ_- to ξ_+. The coarse flow space is the disjoint union of its coarse flow lines $CF_{\xi_-,\xi_+} := \{\xi_-\} \times \Gamma \times \{\xi_+\} \cap CF$. The coarse flow lines are are quasi-isometric to $\mathbb{R}$ (with uniform constants depending on δ).

There are versions of the long thin covers from Remark 3.6 for CF. For $\alpha > 0$ these are VCyc-covers $\mathcal{U}$ of bounded dimension that are α-long in the direction of the coarse flow lines: for $(\xi_-, v, \xi_+) \in CF$ there is $U \in \mathcal{U}$ such that $\{\xi_-\} \times B_\alpha(v) \times \{\xi_+\} \cap CF_{\xi_-,\xi_+} \subseteq U$.

There is also a coarse version of the map φ_t from Remark 3.4. To define it, fix a base point $v_0 \in G$. For $t \in \mathbb{N}$, φ_t sends $\xi \in \partial G$ to (v_0, v, ξ) where $d(v_0, v) = t$ and v belongs to a geodesic from v_0 to ξ. It is convenient to extend φ_t to a map $G \times \partial G \to CF$, with $\varphi_t(g, \xi) := (g v_0, v, \xi)$ where now $d(v_0, v) = t$ and v belongs to a geodesic from $g v_0$ to ξ. Of course v is only coarsely well defined. Nevertheless, φ_t can be used to pull long thin covers for CF back to $G \times \partial G$. For $S \subseteq G$ finite there are then $t > 0$ and $\alpha > 0$ such that this yields S-wide covers for $G \times \partial G$. The proof of this last statement uses a compactness argument and it is important at this point that Γ is locally finite and that G acts cocompactly on Γ.

4 Covers at infinity

The Farrell-Jones Conjecture for $\mathrm{GL}_n(\mathbb{Z})$. The group $\mathrm{GL}_n(\mathbb{Z})$ is not a CAT(0)-group, but it has a proper isometric action on a CAT(0)-space, the symmetric space $X := GL_n(\mathbb{R})/O(n)$. Fix a base point $x_0 \in X$. For $R \geq 0$ let B_R be the closed ball of radius R around x_0. This ball is a retract of X (via the radial projection along geodesics to x_0) and inherits a homotopy coherent action Γ_R from the action of $\mathrm{GL}_n(\mathbb{Z})$ on X. Let $\mathcal{F}$ be the family of subgroups generated by the virtually cyclic and the proper parabolic subgroups of $\mathrm{GL}_n(\mathbb{Z})$. The key step in the proof of the Farrell-Jones Conjecture for $\mathrm{GL}_n(\mathbb{Z})$ in A. Bartels, Lück, Reich, and Rüping [2014] is, in the language of Section 2, the following.

Theorem 4.1. *The sequence of homotopy coherent actions (B_R, Γ_R) is finitely $\mathcal{F}$-amenable.*

In particular $\mathrm{GL}_n(\mathbb{Z})$ satisfies the Farrell-Jones Conjecture relative to $\mathcal{F}$ by Theorem 2.12. Using the transitivity principle 1.2 the Farrell-Jones Conjecture for $\mathrm{GL}_n(\mathbb{Z})$ can then be proven by induction on n. The induction step uses inheritance properties of the Conjecture and that virtually poly-cyclic groups satisfy the Conjecture.

The verification of Theorem 4.1 follows the general strategy of Lemma 3.3 (in a variant for homotopy coherent actions). The additional difficulty in verifying assumption (B) is that, as the action of $\mathrm{GL}_n(\mathbb{Z})$ on the symmetric space is not cocompact, the action on the flow space is not cocompact either. The general results reviewed in Section 3 can still be used to construct for any $\alpha > 0$ an α-long cover $\mathcal{U}$ for the flow space. However it is not clear that the resulting cover is δ-thick, for a $\delta > 0$ uniformly over FS. The remedy for this short-coming is a second collection of open subsets of FS. Its construction starts with an $\mathcal{F}$-cover for X at ∞, meaning here, away from cocompact subsets. Points in the symmetric space can be viewed as inner products on $\mathbb{R}^n$ and moving towards ∞ corresponds to degeneration of inner products along direct summands $W \subset \mathbb{Z}^n \subset \mathbb{R}^n$. This in turn can be used to define horoballs in the symmetric space, one for each W, forming the desired cover Grayson [1984]. For each W the corresponding horoball is invariant for the parabolic subgroup $\{g \in GL_n(\mathbb{Z}) \mid gW = W\}$, more precisely, the horoballs are $\mathcal{F}$-subsets, but not VCyc-subsets. The precise properties of the cover at ∞ are as follows.

Lemma 4.2 (A. Bartels, Lück, Reich, and Rüping [2014] and Grayson [1984])**.** *For any $\alpha > 0$ there exists a collection $\mathcal{U}_\infty$ of open $\mathcal{F}$-subsets of X of order $\leq n$ that is of Lebesgue number $\geq \alpha$ at ∞, i.e., there is $K \subset X$ compact such that for any $x \in X \setminus \mathrm{GL}_n(\mathbb{Z}) \cdot K$ there is $U \in \mathcal{U}_\infty$ containing the α-ball $B_\alpha(x)$ in X around x.*

This cover can be pulled back to the flow space where it provides a cover at ∞ for the flow space that is both (roughly) α-long and α-thick at ∞. Then one is left with a cocompact subset of the flow space where the cover $\mathcal{U}$ constructed first is α-long and δ-thick.

This argument for $GL_n(\mathbb{Z})$ has been generalized to $GL_n(F(t))$ for finite fields F, and $GL_n(\mathbb{Z}[S^{-1}])$, for S a finite set of primes Rüping [2016] using suitable generalizations of the above covers at ∞. In this case the parabolic subgroups are slightly bigger, in particular the induction step (on n) here uses that the Farrell-Jones Conjecture holds for all solvable groups. Using inheritance properties and building on these results the Farrell-Jones Conjecture has been verified for all subgroups of $\mathrm{GL}_n(\mathbb{Q})$ Rüping [ibid.] and all lattices in virtually connected Lie groups Kammeyer, Lück, and Rüping [2016].

Relatively hyperbolic groups. We use Bowditch's characterization of relatively hyperbolic groups Bowditch [2012]. A graph is *fine* if there are only finitely many embedded loops of a given length containing a given edge. Let $\mathcal{P}$ be a collection of subgroups of the countable group G. Then G is hyperbolic relative to $\mathcal{P}$ if G admits a cocompact action on a fine hyperbolic graph Γ such that all edge stabilizers are finite and all vertex stabilizers belong to $\mathcal{P}$. The subgroups from $\mathcal{P}$ are said to be peripheral or parabolic. The requirement that Γ is fine encodes Farb's Bounded Coset Penetration property Farb [1998]. Bowditch assigned a compact boundary Δ to G as follows. As a set Δ is the union of the Gromov boundary $\partial\Gamma$ with the set of all vertices of infinite valency in Γ. The topology is the observer topology; a sequence x_n converges in this topology to x if given any finite set S of vertices (not including x), for almost all n there is a geodesic from x_n to x that misses S. (For general hyperbolic graphs this topology is not Hausdorff, but for fine hyperbolic graphs it is.)

The main result from A. Bartels [2017] is that if G is hyperbolic relative to $\mathcal{P}$, then G satisfies the Farrell-Jones Conjecture relative to the family of subgroups $\mathcal{F}$ generated by VCyc and $\mathcal{P}$ ($\mathcal{P}$ needs to be closed under index two supergroups here for this to include the L-theoretic version of the Farrell-Jones Conjecture). This result is obtained as an application of Theorem 2.6. The key step is the following.

Theorem 4.3 (A. Bartels [ibid.]). *The action of G on Δ is finitely $\mathfrak{F}$-amenable.*

This is a direct consequence of Propositions 4.4 and 4.5 below, using the characterization of N-$\mathfrak{F}$-amenability from Remark 2.13 by the existence of S-wide covers of $G \times \Delta$.

To outline the construction of these covers and to prepare for the mapping class group we introduce some notation. Pick a G-invariant proper metric on the set E of edges of Γ; this is possible as G and E are countable and the action of G on E has finite stabilizers. For each vertex v of Γ with infinite valency let E_v be the set of edges incident to v. Write d_v for the restriction of the metric to E_v. For $\xi \in \Delta$, $\xi \neq v$ we define its projection $\pi_v(\xi)$ to E_v as the set of all edges of Γ that are appear as initial edges of geodesics from v to ξ. This is a finite subset of E_v (this depends again of fineness of Γ). Fix a vertex v_0 of finite valence as a base point. For $g \in G$, $\xi \in \Delta$ define their *projection distance* at v by

$$d_v^{\pi}(g,\xi) := d_v(\pi_v(g v_0), \pi_v(\xi)).$$

For $\xi = v$, set $d_v^{\pi}(g, v) := \infty$. (For relative hyperbolic groups a related quantity is often called an *angle*; the terminology here is chosen to align better with the case of the mapping class group.) If we vary g (in a finite set) and ξ (in an open neighborhood) then for fixed v the projection distance $d_v^{\pi}(g,\xi)$ varies by a bounded amount. Useful is the following attraction property for projection distances: there is Θ_0 such that if $d_v^{\pi}(g,\xi) \geq \Theta_0$, then any geodesic from $g v_0$ to ξ in Γ passes through v. Conversely, if some geodesic from $g v_0$ to ξ misses v, then $d_v^{\pi}(g,\xi) < \Theta_0'$ for some uniform Θ_0'.

Projection distances are used to control the failure of Γ to be locally finite. In particular, provided all projection distances are bounded by a constant Θ, a variation of the argument for hyperbolic groups (using a coarse flow space), can be adapted to provide S-long covers for the Θ-small part of $G \times \Delta$. The following is a precise statement.

Proposition 4.4. *There is N (depending only on G and Δ) such that for any $\Theta > 0$ and any $S \subseteq G$ finite there exists a collection $\mathcal{U}$ of open* VCyc*-subsets of $G \times \Delta$ that is S-wide on the Θ-small part, i.e., if $(g,\xi) \in G \times \Delta$ satisfies $d_v^{\pi}(g,\xi) \leq \Theta$ for all vertices v, then there is $U \in \mathcal{U}$ with $gS \times \{\xi\} \subseteq U$.*

To deal with large projection distances an explicit construction can be used (similar to the case of $\mathrm{GL}_n(\mathbb{Z})$). For $(g,\xi) \in G \times \Delta$ let

$$V_{\Theta}(g,\xi) := \{v \mid d_v^{\pi}(g,\xi) \geq \Theta\}.$$

As a consequence of the attraction property, for sufficiently large Θ, the set $V_{\Theta}(g,\xi)$ consists of vertices that belong to any geodesic from $g v_0$ to ξ. In particular, it can be linearly ordered by distance from $g v_0$.

For a fixed vertex v and $\Theta > 0$ define $W(v,\Theta) \subset G \times \Delta$ as the (interior of the) set of all pairs (g,ξ) for which v is minimal in $V_{\Theta}(g,\xi)$, i.e., v is the vertex closest to $g v_0$

for which $d^\pi_v(g,\xi) \geq \Theta$. Then $\mathcal{W}(\Theta) := \{W(v,\Theta) \mid v \in V\}$ is a collection of pairwise disjoint open $\mathcal{P}$-subsets of $G\times\Delta$.

Proposition 4.5. *Let $S \subset G$ be finite. Then there are $\theta'' \gg \theta' \gg \theta \gg 0$ such that $\mathcal{W}(\theta) \cup \mathcal{W}(\theta')$ is a G-invariant collection of open $\mathcal{P}$-subsets of order ≤ 1 that is S-long on the θ''-large part of $G\times\Delta$: if $d^\pi_v(g,\xi) \geq \theta''$ for some vertex v, then there is $W \in \mathcal{W}(\theta) \cup \mathcal{W}(\theta')$ such that $gS\times\{\xi\} \subset W$.*

A difficulty in working with the $\mathcal{W}(v,\Theta)$ is that it is for fixed Θ not possible to control exactly how $V_\Theta(g,\xi)$ varies with g and ξ. In particular whether or not a vertex v is minimal in $V_\Theta(g,\xi)$ can change under small variation in g or ξ. A consequence of the attraction property that is useful for the proof of Proposition 4.5 is the following: suppose there are vertices v_0 and v_1 with $d^\pi_{v_i}(g,\xi) > \Theta' \gg \Theta$, then the segment between v_0 and v_1 in the linear order of $V_\Theta(g,\xi)$ is unchanged under suitable variations of (g,ξ) depending on Θ'.

Remark 4.6. A motivating example of relatively hyperbolic groups are fundamental groups G of complete Riemannian manifolds M of pinched negative sectional curvature and finite volume. These are hyperbolic relative to their virtually finitely generated nilpotent subgroups Bowditch [2012] and Farb [1998]. In this case we can work with the sphere at ∞ of the universal cover $\tilde{M}$ of M. The splitting of $G\times S_\infty$ into a Θ-small part and a Θ-large part can be thought of as follows. Fix a base point $x_0 \in \tilde{M}$. Instead of a number Θ we choose a cocompact subset $G \cdot K$ of $\tilde{M}$. The small part of $G\times\Delta$ consists then of all pairs (g,ξ) for which the geodesic ray from gx_0 to ξ is contained in X; the large part is the complement. Under this translation the cover from Proposition 4.5 can again be thought of as a cover at ∞ for $\tilde{M}$. Moreover, the vertices of infinite valency in Γ correspond to horoballs in $\tilde{M}$, and projection distances to time geodesic rays spend in horoballs.

Note that the action of G on the graph Γ in the definition of relative hyperbolicity we used is cocompact, but Γ is not a proper metric space. Conversely, in the above example the action of G on $\tilde{M}$ is no longer cocompact, but now $\tilde{M}$ is a proper metric space. A similar trade off (cocompact action on non proper space versus non-cocompact action on proper space) is possible for all relatively hyperbolic groups Groves and Manning [2008], assuming the parabolic subgroups are finitely generated.

The mapping class group. Let Σ be a closed orientable surface of genus g with a finite set P of p marked points. We will assume $6g + 2p - 6 > 0$. The mapping class group $\mathrm{Mod}(\Sigma)$ of Σ is the group of components of the group of orientation preserving homeomorphisms of Σ that leave P invariant. Teichmüller space $\mathcal{T}$ is the space of marked complete hyperbolic structures of finite area on $\Sigma \setminus P$. The mapping class group acts on Teichmüller space by changing the marking. Thurston defined an equivariant compactification of Teichmüller space $\overline{\mathcal{T}}$, see Fathi, Laudenbach, and Poénaru [2012]. As a space $\overline{\mathcal{T}}$

is a closed disk, in particular it is an Euclidean retract. The boundary of the compactification $\mathcal{PMF} := \overline{\mathcal{T}} \setminus \mathcal{T}$ is the space of projective measured foliations on Σ. The key step in the proof of the Farrell-Jones Conjecture for $\mathrm{Mod}(\Sigma)$ is the following.

Theorem 4.7 (A. Bartels and Bestvina [2016]). *Let $\mathcal{F}$ be the family of subgroups of* $\mathrm{Mod}(\Sigma)$ *that virtually fix a point in $\mathcal{PMF}$. The action of* $\mathrm{Mod}(\Sigma)$ *on $\mathcal{PMF}$ is finitely $\mathcal{F}$-amenable.*

From this it follows quickly that the action on $\overline{\mathcal{T}}$ is finitely $\mathcal{F}$-amenable as well, and applying Theorem 2.6 we obtain the Farrell-Jones Conjecture for $\mathrm{Mod}(\Sigma)$ relative to $\mathcal{F}$. Up to passing to finite index subgroups, the groups in $\mathcal{F}$ are central extensions of products of mapping class groups of smaller complexity. Using the transitivity principle and inheritance properties one then obtains the Farrell-Jones Conjecture for $\mathrm{Mod}(\Sigma)$ by induction on the complexity of Σ. The only additional input in this case is that the Farrell-Jones Conjecture holds for finitely generated free abelian groups.

The proof of Theorem 4.7 uses the characterization of N-$\mathcal{F}$-amenability from Remark 2.13 and provides suitable covers for $\mathrm{Mod}(\Sigma) \times \mathcal{PMF}$. Similar to the relative hyperbolic case the construction of these covers is done by splitting $\mathrm{Mod}(\Sigma) \times \mathcal{PMF}$ into two parts. Here it is natural to refer to these parts as the thick part and the thin part. (The thick part corresponds to the Θ-small part in the relative hyperbolic case.)

Teichmüller space has a natural filtration by cocompact subsets. For $\varepsilon > 0$ the ε-thick part $\mathcal{T}_{\geq\varepsilon} \subseteq \mathcal{T}$ consist of all marked hyperbolic structures such that all closed geodesics have length $\geq \varepsilon$. The action of $\mathrm{Mod}(\Sigma)$ on $\mathcal{T}_{\geq\varepsilon}$ is cocompact Mumford [1971]. Fix a base point $x_0 \in \mathcal{T}$. Given a pair $(g, \xi) \in \mathrm{Mod}(\Sigma) \times \mathcal{PMF}$ there is a unique Teichmüller ray c_g, that starts at $g(x_0)$ and is "pointing towards ξ" (technically, the vertical foliation of the quadratic differential is ξ). The ε-thick part of $\mathrm{Mod}(\Sigma) \times \mathcal{PMF}$ is defined as the set of all pairs (g, ξ) for which the Teichmüller ray $c_{g,\xi}$ stays in $\mathcal{T}_{\geq\varepsilon}$.

An important tool in covering both the thick and the thin part is the complex of curves $\mathcal{C}(\Sigma)$. A celebrated result of Mazur-Minsky is that $\mathcal{C}(\Sigma)$ is hyperbolic H. A. Masur and Minsky [1999]. Klarreich [1999] studied a coarse projection map $\pi \colon \mathcal{T} \to \mathcal{C}(\Sigma)$ and identified the Gromov boundary $\partial\mathcal{C}(\Sigma)$ of the curve complex. In particular, the projection map has an extension $\pi \colon \mathcal{PMF} \to \mathcal{C}(\Sigma) \cup \partial\mathcal{C}(\Sigma)$. (On the preimage of the Gromov boundary this extension is a continuous map; on the complement it is still only a coarse map.)

Teichmüller space is not hyperbolic, but its thick part $\mathcal{T}_{\geq\varepsilon}$ has a number of hyperbolic properties: The Masur criterion H. Masur [1992] implies that for (g, ξ) in the thick part, $c_{g,\xi}(t) \to \xi$ as $t \to \infty$. Moreover, the restriction of Klarreichs projection map $\pi \colon \mathcal{PMF} \to \mathcal{C}(\Sigma) \cup \partial\mathcal{C}(\Sigma)$ to the space of all such ξ is injective. A result of Minsky [1996] is that geodesics c that stay in $\mathcal{T}_{\geq\varepsilon}$ are contracting. This is a property they share with geodesics in hyperbolic spaces: the nearest point projection $\mathcal{T} \to c$ maps balls disjoint from c to uniformly bounded subsets. Teichmüller geodesics in the thick part $\mathcal{T}_{\geq\varepsilon}$ project to quasi-geodesics in the curve complex with constants depending only on ε. All these properties eventually allow for the construction of suitable covers of any thick part using a coarse flow space and methods from the hyperbolic case. A precise statement is the following.

Proposition 4.8. *There is d such that for any $\varepsilon > 0$ and any $S \subset \mathrm{Mod}(\Sigma)$ finite there exists a $\mathrm{Mod}(\Sigma)$-invariant collection $\mathcal{U}$ of $\mathcal{F}$-subsets of $\mathrm{Mod}(\Sigma)\times\mathcal{PMF}$ of order $\leq d$ such for any (g,ξ) for which $c_{g,\xi}$ stays in $\mathcal{T}_{\geq\varepsilon}$ there is $U \in \mathcal{U}$ with $gS\times\{\xi\} \subseteq U$.*

The action of the mapping class group on the curve complex $\mathcal{C}(\Sigma)$ does not exhibit the mapping class group as a relative hyperbolic group in the sense discussed before; the 1-skeleton of $\mathcal{C}(\Sigma)$ is not fine. Nevertheless, there is an important replacement for the projections to links used in the relatively hyperbolic case, the subsurface projections of H. A. Masur and Minsky [2000]. In this case the projections are not to links in the curve complex, but to curve complexes $\mathcal{C}(Y)$ of subsurfaces Y of Σ. (On the other hand, often links in the curve complex are exactly curve complexes of subsurfaces.) The theory is however much more sophisticated than in the relatively hyperbolic case. Projections are not always defined; sometimes the projection is to points in the boundary of $\mathcal{C}(Y)$ and the projection distance is ∞. Bestvina, Bromberg, and Fujiwara [2015] used subsurface projections to prove that the mapping class group has finite asymptotic dimension. In their work the subsurfaces of Σ are organized in a finite number N of families $\mathbf{Y}$ such that two subsurfaces in the same family will always intersect in an interesting way. This has the effect that the projections for subsurfaces in the same family interact in a controlled way with each other. Each family $\mathbf{Y}$ of subsurfaces is organized in Bestvina, Bromberg, and Fujiwara [ibid.] in an associated simplicial complex, called the projection complex. The vertices of the projection complex are the subsurfaces from $\mathbf{Y}$. A perturbation of the projection distances can thought of as being measured along geodesics in the projection complex and now behaves very similar as in the relative hyperbolic case, in particular the attraction property is satisfied in each projection complex. This allows the application of a variant of the construction from Proposition 4.5 for each projection complex that eventually yield the following.

Proposition 4.9. *Let $\mathbf{Y}$ be any of the finitely many families of subsurfaces. For any $S \subseteq G$ finite there exists $\Theta > 0$ and a $\mathrm{Mod}(\Sigma)$-invariant collection $\mathcal{U}$ of $\mathcal{F}$-subsets of $\mathrm{Mod}(\Sigma)\times\mathcal{PMF}$ of order ≤ 1 such for any (g,ξ) for which there is $Y \in \mathbf{Y}$ with $d^{\pi}_{Y}(g,\xi) \geq \Theta$ there is $U \in \mathcal{U}$ with $gS\times\{\xi\} \subseteq U$.*

The final piece, needed to combine Propositions 4.8 and 4.9 to a proof of Theorem 4.7, is a consequence of Rafi's analysis of short curves in Σ along Teichmüller rays Rafi [2005]:

for any $\varepsilon > 0$ there is Θ such that if some curve of Σ is ε-short on $c_{g,\xi}$ (i.e., if $c_{g,\xi}$ is not contained in $\mathfrak{T}_{\geq\varepsilon}$) then there is a subsurface Y such that $d^{\pi}_{Y}(g,\xi) \geq \Theta$.

Remark 4.10. Farrell and Jones [1998] proved topological rigidity results for fundamental groups of non-positively curved manifolds that are in addition A-regular. The latter condition bounds the curvature tensor and its covariant derivatives over the manifold. All torsion free discrete subgroups of $\mathrm{GL}_n(\mathbb{R})$ are fundamental groups of such manifolds. Similar to the examples discussed in this section a key difficulty in Farrell and Jones [ibid.] is that the action of the fundamental group G of the manifold on the universal cover is not cocompact. The general strategy employed by Farrell-Jones seems however different, in particular, it does not involve an induction over some kind of complexity of G. The only groups that are considered in an intermediate step are polycyclic groups, and the argument directly reduces from G to these and then uses computations of K-and L-theory for polycyclic groups.

This raises the following question: Can the family of subgroups in Theorem 4.1 be replaced with the family of virtually polycyclic subgroups? Recall that the cover constructed for the flow space at ∞ is both α-long and α-thick, while only δ-thickness is needed. So it is plausible that there exist thinner covers at ∞ that work for the family of virtually polycyclic subgroups.

For the mapping class group the family from Theorem 4.7 can not be chosen to be significantly smaller; all isotropy groups for the action have to appear in the family. But one might ask, whether there exist N-$\mathfrak{F}$-amenable actions (or N-$\mathfrak{F}$-amenable sequences of homotopy coherent actions) of mapping class groups on Euclidean retracts, where $\mathfrak{F}$ is smaller than the family used in Theorem 4.7.

References

S. Adams (1994). "Boundary amenability for word hyperbolic groups and an application to smooth dynamics of simple groups". *Topology* 33.4, pp. 765–783. MR: 1293309 (cit. on p. 1063).

Goulnara Arzhantseva, Martin R. Bridson, Tadeusz Januszkiewicz, Ian J. Leary, Ashot Minasyan, and Jacek Świ tkowski (2009). "Infinite groups with fixed point properties". *Geom. Topol.* 13.3, pp. 1229–1263. MR: 2496045 (cit. on p. 1065).

A. Bartels (2017). "Coarse flow spaces for relatively hyperbolic groups". *Compos. Math.* 153.4, pp. 745–779. MR: 3631229 (cit. on pp. 1070, 1072, 1073).

A. Bartels and W. Lück (2012a). "The Farrell-Hsiang method revisited". *Math. Ann.* 354.1, pp. 209–226. MR: 2957625 (cit. on p. 1067).

Arthur Bartels (2016). "On proofs of the Farrell-Jones conjecture". In: *Topology and geometric group theory*. Vol. 184. Springer Proc. Math. Stat. Springer, [Cham], pp. 1–31. MR: 3598160 (cit. on p. 1065).

Arthur C. Bartels (2003). "Squeezing and higher algebraic K-theory". *K-Theory* 28.1, pp. 19–37. MR: 1988817 (cit. on p. 1064).

Arthur Bartels and Mladen Bestvina (2016). "The Farrell-Jones Conjecture for mapping class groups". Preprint. arXiv: 1606.02844 (cit. on p. 1075).

Arthur Bartels, F. T. Farrell, L. E. Jones, and Holger Reich (2004). "On the isomorphism conjecture in algebraic K-theory". *Topology* 43.1, pp. 157–213. MR: 2030590 (cit. on p. 1061).

Arthur Bartels and Wolfgang Lück (2012b). "Geodesic flow for $\mathrm{CAT}(0)$-groups". *Geom. Topol.* 16.3, pp. 1345–1391. MR: 2967054 (cit. on pp. 1067, 1068).

– (2012c). "The Borel conjecture for hyperbolic and $\mathrm{CAT}(0)$-groups". *Ann. of Math. (2)* 175.2, pp. 631–689. MR: 2993750 (cit. on pp. 1060, 1064, 1066, 1067).

Arthur Bartels, Wolfgang Lück, and Holger Reich (2008a). "Equivariant covers for hyperbolic groups". *Geom. Topol.* 12.3, pp. 1799–1882. MR: 2421141 (cit. on pp. 1065, 1070).

– (2008b). "The K-theoretic Farrell-Jones conjecture for hyperbolic groups". *Invent. Math.* 172.1, pp. 29–70. MR: 2385666 (cit. on pp. 1060, 1064, 1067).

Arthur Bartels, Wolfgang Lück, Holger Reich, and Henrik Rüping (2014). "K- and L-theory of group rings over $GL_n(\mathbf{Z})$". *Publ. Math. Inst. Hautes Études Sci.* 119, pp. 97–125. MR: 3210177 (cit. on pp. 1071, 1072).

Arthur Bartels and Holger Reich (2007). "Coefficients for the Farrell-Jones conjecture". *Adv. Math.* 209.1, pp. 337–362. MR: 2294225 (cit. on p. 1062).

Paul Baum and Alain Connes (2000). "Geometric K-theory for Lie groups and foliations". *Enseign. Math. (2)* 46.1-2, pp. 3–42. MR: 1769535 (cit. on p. 1060).

Paul Baum, Alain Connes, and Nigel Higson (1994). "Classifying space for proper actions and K-theory of group C^*-algebras". In: *C^*-algebras: 1943–1993 (San Antonio, TX, 1993)*. Vol. 167. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 240–291. MR: 1292018 (cit. on p. 1060).

Mladen Bestvina, Ken Bromberg, and Koji Fujiwara (2015). "Constructing group actions on quasi-trees and applications to mapping class groups". *Publ. Math. Inst. Hautes Études Sci.* 122, pp. 1–64. MR: 3415065 (cit. on p. 1076).

Mladen Bestvina, Vincent Guirardel, and Camille Horbez (2017). "Boundary amenability of $Out(F_N)$". Preprint. arXiv: 1705.07017 (cit. on p. 1063).

Mladen Bestvina and Geoffrey Mess (1991). "The boundary of negatively curved groups". *J. Amer. Math. Soc.* 4.3, pp. 469–481. MR: 1096169 (cit. on p. 1065).

M. Bökstedt, W. C. Hsiang, and I. Madsen (1993). "The cyclotomic trace and algebraic K-theory of spaces". *Invent. Math.* 111.3, pp. 465–539. MR: 1202133 (cit. on p. 1068).

B. H. Bowditch (2012). "Relatively hyperbolic groups". *Internat. J. Algebra Comput.* 22.3, pp. 1250016, 66. MR: 2922380 (cit. on pp. 1072, 1074).

Gunnar Carlsson and Boris Goldfarb (2004). "The integral K-theoretic Novikov conjecture for groups with finite asymptotic dimension". *Invent. Math.* 157.2, pp. 405–418. MR: 2076928 (cit. on p. 1064).

James F. Davis and Wolfgang Lück (1998). "Spaces over a category and assembly maps in isomorphism conjectures in K- and L-theory". *K-Theory* 15.3, pp. 201–252. MR: 1659969 (cit. on p. 1061).

Andreas W. M. Dress (1975). "Induction and structure theorems for orthogonal representations of finite groups". *Ann. of Math. (2)* 102.2, pp. 291–325. MR: 0387392 (cit. on p. 1067).

Nils-Edvin Enkelmann, Wolfgang Lück, Malte Pieper, Mark Ullmann, and Christoph Winges (2016). "On the Farrell-Jones Conjecture for Waldhausen's A-theory". Preprint. arXiv: 1607.06395v2 (cit. on pp. 1062, 1067).

B. Farb (1998). "Relatively hyperbolic groups". *Geom. Funct. Anal.* 8.5, pp. 810–840. MR: 1650094 (cit. on pp. 1072, 1074).

F. T. Farrell and W. C. Hsiang (1978). "The topological-Euclidean space form problem". *Invent. Math.* 45.2, pp. 181–192. MR: 0482771 (cit. on p. 1067).

F. T. Farrell and L. E. Jones (1986). "K-theory and dynamics. I". *Ann. of Math. (2)* 124.3, pp. 531–569. MR: 866708 (cit. on pp. 1059, 1066, 1070).

– (1993a). "Isomorphism conjectures in algebraic K-theory". *J. Amer. Math. Soc.* 6.2, pp. 249–297. MR: 1179537 (cit. on pp. 1060–1062).

– (1993b). "Topological rigidity for compact non-positively curved manifolds". In: *Differential geometry: Riemannian geometry (Los Angeles, CA, 1990)*. Vol. 54. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 229–274. MR: 1216623 (cit. on pp. 1060, 1066, 1069).

– (1998). "Rigidity for aspherical manifolds with $\pi_1 \subset \mathrm{GL}_m(\mathbf{R})$". *Asian J. Math.* 2.2, pp. 215–262. MR: 1639544 (cit. on p. 1077).

Albert Fathi, François Laudenbach, and Valentin Poénaru (2012). *Thurston's work on surfaces*. Vol. 48. Mathematical Notes. Translated from the 1979 French original by Djun M. Kim and Dan Margalit. Princeton University Press, Princeton, NJ, pp. xvi+254. MR: 3053012 (cit. on p. 1074).

Daniel R. Grayson (1984). "Reduction theory using semistability". *Comment. Math. Helv.* 59.4, pp. 600–634. MR: 780079 (cit. on p. 1072).

Daniel Groves and Jason Fox Manning (2008). "Dehn filling in relatively hyperbolic groups". *Israel J. Math.* 168, pp. 317–429. MR: 2448064 (cit. on p. 1074).

Erik Guentner, Nigel Higson, and Shmuel Weinberger (2005). "The Novikov conjecture for linear groups". *Publ. Math. Inst. Hautes Études Sci.* 101, pp. 243–268. MR: 2217050 (cit. on p. 1063).

Erik Guentner, Romain Tessera, and Guoliang Yu (2012). "A notion of geometric complexity and its application to topological rigidity". *Invent. Math.* 189.2, pp. 315–357. MR: 2947546 (cit. on p. 1064).

Erik Guentner, Rufus Willett, and Guoliang Yu (2017). "Dynamic asymptotic dimension: relation to dynamics, topology, coarse geometry, and C^*-algebras". *Math. Ann.* 367.1-2, pp. 785–829. MR: 3606454 (cit. on pp. 1064, 1067).

Ian Hambleton and Erik K. Pedersen (2004). "Identifying assembly maps in K- and L-theory". *Math. Ann.* 328.1-2, pp. 27–57. MR: 2030369 (cit. on p. 1061).

Ursula Hamenstädt (2009). "Geometry of the mapping class groups. I. Boundary amenability". *Invent. Math.* 175.3, pp. 545–609. MR: 2471596 (cit. on p. 1063).

N. Higson (2000). "Bivariant K-theory and the Novikov conjecture". *Geom. Funct. Anal.* 10.3, pp. 563–581. MR: 1779613 (cit. on p. 1063).

Nigel Higson and Gennadi Kasparov (2001). "E-theory and KK-theory for groups which act properly and isometrically on Hilbert space". *Invent. Math.* 144.1, pp. 23–74. MR: 1821144 (cit. on p. 1060).

Nigel Higson and John Roe (2000). "Amenable group actions and the Novikov conjecture". *J. Reine Angew. Math.* 519, pp. 143–153. MR: 1739727 (cit. on p. 1064).

Wu Chung Hsiang (1984). "Geometric applications of algebraic K-theory". In: *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*. PWN, Warsaw, pp. 99–118. MR: 804679 (cit. on p. 1059).

Holger Kammeyer, Wolfgang Lück, and Henrik Rüping (2016). "The Farrell-Jones conjecture for arbitrary lattices in virtually connected Lie groups". *Geom. Topol.* 20.3, pp. 1275–1287. MR: 3523058 (cit. on pp. 1060, 1072).

D. Kasprowski, M. Ullmann, C. Wegner, and C. Winges (2018). "The A-theoretic Farrell-Jones conjecture for virtually solvable groups". *Bull. London Math. Soc.* 50.2, pp. 219–228. arXiv: 1609.07468 (cit. on p. 1062).

Daniel Kasprowski (2015). "On the K-theory of groups with finite decomposition complexity". *Proc. Lond. Math. Soc. (3)* 110.3, pp. 565–592. MR: 3342098 (cit. on p. 1064).

Daniel Kasprowski and Henrik Rüping (2017). "Long and thin covers for flow spaces". *Groups Geom. Dyn.* 11.4, pp. 1201–1229. arXiv: 1502.05001. MR: 3737280 (cit. on p. 1070).

Yoshikata Kida (2008). "The mapping class group from the viewpoint of measure equivalence theory". *Mem. Amer. Math. Soc.* 196.916, pp. viii+190. MR: 2458794 (cit. on p. 1063).

Erica Klarreich (1999). "The boundary at infinity of the complex of curves and the relative teichmüller space". arXiv: 1803.10339 (cit. on p. 1075).

Svenja Knopf (2017). "Acylindrical Actions on Trees and the Farrell-Jones Conjecture". Preprint. arXiv: 1704.05671 (cit. on p. 1065).

Vincent Lafforgue (2002). "K-théorie bivariante pour les algèbres de Banach et conjecture de Baum-Connes". *Invent. Math.* 149.1, pp. 1–95. MR: 1914617 (cit. on p. 1060).

– (2012). "La conjecture de Baum-Connes à coefficients pour les groupes hyperboliques". *J. Noncommut. Geom.* 6.1, pp. 1–197. MR: 2874956 (cit. on p. 1060).

Wolfgang Lück (2005). "Survey on classifying spaces for families of subgroups". In: *Infinite groups: geometric, combinatorial and dynamical aspects*. Vol. 248. Progr. Math. Birkhäuser, Basel, pp. 269–322. MR: 2195456 (cit. on p. 1061).

– (2010). "K- and L-theory of group rings". In: *Proceedings of the International Congress of Mathematicians. Volume II*. Hindustan Book Agency, New Delhi, pp. 1071–1098. MR: 2827832 (cit. on p. 1060).

– (2017). *Slides for a talk in oberwolfach: The Farrell-Jones Conjecture* (cit. on p. 1061).

Wolfgang Lück and Holger Reich (2005). "The Baum-Connes and the Farrell-Jones conjectures in K- and L-theory". In: *Handbook of K-theory. Vol. 1, 2*. Springer, Berlin, pp. 703–842. MR: 2181833 (cit. on p. 1062).

Wolfgang Lück, Holger Reich, John Rognes, and Marco Varisco (2017). "Algebraic K-theory of group rings and the cyclotomic trace map". *Adv. Math.* 304, pp. 930–1020. MR: 3558224 (cit. on p. 1068).

H. A. Masur and Y. N. Minsky (2000). "Geometry of the complex of curves. II. Hierarchical structure". *Geom. Funct. Anal.* 10.4, pp. 902–974. MR: 1791145 (cit. on p. 1076).

Howard Masur (1992). "Hausdorff dimension of the set of nonergodic foliations of a quadratic differential". *Duke Math. J.* 66.3, pp. 387–442. MR: 1167101 (cit. on p. 1075).

Howard A. Masur and Yair N. Minsky (1999). "Geometry of the complex of curves. I. Hyperbolicity". *Invent. Math.* 138.1, pp. 103–149. MR: 1714338 (cit. on p. 1075).

Igor Mineyev (2005). "Flows and joins of metric spaces". *Geom. Topol.* 9, pp. 403–482. MR: 2140987 (cit. on p. 1068).

Igor Mineyev and Guoliang Yu (2002). "The Baum-Connes conjecture for hyperbolic groups". *Invent. Math.* 149.1, pp. 97–122. MR: 1914618 (cit. on p. 1060).

Yair N. Minsky (1996). "Quasi-projections in Teichmüller space". *J. Reine Angew. Math.* 473, pp. 121–136. MR: 1390685 (cit. on p. 1075).

David Mumford (1971). "A remark on Mahler's compactness theorem". *Proc. Amer. Math. Soc.* 28, pp. 289–294. MR: 0276410 (cit. on p. 1075).

Robert Oliver (1975). "Fixed-point sets of group actions on finite acyclic complexes". *Comment. Math. Helv.* 50, pp. 155–177. MR: 0375361 (cit. on p. 1065).

Narutaka Ozawa (2006). "Amenable actions and applications". In: *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, pp. 1563–1580. MR: 2275659 (cit. on p. 1063).

Erik Kjær Pedersen (2000). "Controlled algebraic K-theory, a survey". In: *Geometry and topology: Aarhus (1998)*. Vol. 258. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 351–368. MR: 1778117 (cit. on p. 1065).

Kasra Rafi (2005). "A characterization of short curves of a Teichmüller geodesic". *Geom. Topol.* 9, pp. 179–202. MR: 2115672 (cit. on p. 1076).

Daniel A. Ramras, Romain Tessera, and Guoliang Yu (2014). "Finite decomposition complexity and the integral Novikov conjecture for higher algebraic K-theory". *J. Reine Angew. Math.* 694, pp. 129–178. MR: 3259041 (cit. on p. 1064).

Holger Reich and Marco Varisco (2017). "Algebraic K-theory, assembly maps, controlled algebra, and trace methods". Preprint. arXiv: 1702.02218 (cit. on pp. 1061, 1062).

H. Rüping (2016). "The Farrell-Jones conjecture for S-arithmetic groups". *J. Topol.* 9.1, pp. 51–90. MR: 3465840 (cit. on p. 1072).

Damian Sawicki (2017). "On equivariant asymptotic dimension". *Groups Geom. Dyn.* 11.3, pp. 977–1002. MR: 3692903 (cit. on pp. 1067, 1070).

Richard G. Swan (1960). "Induced representations and projective modules". *Ann. of Math. (2)* 71, pp. 552–578. MR: 0138688 (cit. on p. 1067).

Mark Ullmann and Christoph Winges (2015). "On the Farrell-Jones Conjecture for algebraic K-theory of spaces: the Farrell-Hsiang method". Preprint. arXiv: 1509.07363 (cit. on p. 1062).

Rainer M. Vogt (1973). "Homotopy limits and colimits". *Math. Z.* 134, pp. 11–52. MR: 0331376 (cit. on p. 1066).

Christian Wegner (2012). "The K-theoretic Farrell-Jones conjecture for CAT(0)-groups". *Proc. Amer. Math. Soc.* 140.3, pp. 779–793. MR: 2869063 (cit. on pp. 1066, 1067).

– (2015). "The Farrell-Jones conjecture for virtually solvable groups". *J. Topol.* 8.4, pp. 975–1016. MR: 3431666 (cit. on pp. 1060, 1067).

Guoliang Yu (1998). "The Novikov conjecture for groups with finite asymptotic dimension". *Ann. of Math. (2)* 147.2, pp. 325–355. MR: 1626745 (cit. on p. 1064).

– (2000). "The coarse Baum-Connes conjecture for spaces which admit a uniform embedding into Hilbert space". *Invent. Math.* 139.1, pp. 201–240. MR: 1728880 (cit. on p. 1063).

– (2017). "The Novikov conjecture for algebraic K-theory of the group algebra over the ring of Schatten class operators". *Adv. Math.* 307, pp. 727–753. MR: 3590530 (cit. on p. 1068).

Received 2017-11-29.

Arthur Bartels
WWU Münster
Mathematisches Institut
Germany
a.bartels@wwu.de

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1081–1104)

KNOT CONTACT HOMOLOGY AND OPEN GROMOV–WITTEN THEORY

Tobias Ekholm

Abstract

Knot contact homology studies symplectic and contact geometric properties of conormals of knots in 3-manifolds using holomorphic curve techniques. It has connections to both mathematical and physical theories. On the mathematical side, we review the theory, show that it gives a complete knot invariant, and discuss its connections to Fukaya categories, string topology, and micro-local sheaves. On the physical side, we describe the connection between the augmentation variety of knot contact homology and Gromov–Witten disk potentials, and discuss the corresponding higher genus relation that quantizes the augmentation variety.

1 Introduction

If M is an oriented 3-manifold then its 6-dimensional cotangent bundle T^*M with the closed non-degenerate 2-form $\omega = -d\theta$, where $\theta = p\,dq$ is the Liouville or action 1-form, is a symplectic manifold. As a symplectic manifold, T^*M satisfies the Calabi–Yau condition, $c_1(T^*M) = 0$, and is thus a natural ambient space for the topological string theory of physics and its mathematical counterpart, Gromov–Witten theory.

If $K \subset M$ is a knot then its *Lagrangian conormal* $L_K \subset T^*M$ of covectors along K that annihilate the tangent vector of K is a Lagrangian submanifold (i.e., $\omega|_{L_K} = 0$) diffeomorphic to $S^1 \times \mathbb{R}^2$. Lagrangian submanifolds provide natural boundary conditions for open string theory or open Gromov–Witten theory, that counts holomorphic curves with boundary on the Lagrangian.

Here we will approach the Gromov–Witten theory of L_K from geometric data at infinity. At infinity, the pair (T^*M, L_K) has ideal contact boundary (ST^*M, Λ_K), the unit sphere cotangent bundle ST^*M with the contact form $\alpha = \theta|_{ST^*M}$ and Λ_K the *Legendrian conormal* ($\alpha|_{\Lambda_K} = 0$) $\Lambda_K = L_K \cap ST^*M$. In what follows we will restrict attention to the most basic cases of knots in 3-space or the 3-sphere, $M = \mathbb{R}^3$ or $M = S^3$.

The author is supported by the Knut and Alice Wallenberg Foundation and by the Swedish Research Council.
MSC2010: primary 53D42; secondary 53D37, 53D45, 57R17, 57M25.

1.1 Mathematical aspects of knot contact homology. There is a variety of holomorphic curve theories, all interconnected, that can be applied to distinguish objects up to deformation in contact and symplectic geometry. Knot contact homology belongs to a framework of such theories called Symplectic Field Theory (SFT) Eliashberg, Givental, and Hofer [2000]. More precisely, it is the most basic version of SFT, the Chekanov–Eliashberg dg-algebra $CE(\Lambda_K)$, of the Legendrian conormal torus $\Lambda_K \subset ST^*\mathbb{R}^3$ of a knot $K \subset \mathbb{R}^3$. The study of knot contact homology was initiated by Eliashberg, see Eliashberg [2007], around 2000 and developed from a combinatorial perspective by Ng [2008, 2011] and with holomorphic curve techniques in Ekholm, J. B. Etnyre, Ng, and M. G. Sullivan [2013] and Ekholm, J. Etnyre, Ng, and M. Sullivan [2013].

Our first result states that the contact deformation class of Λ_K encodes the isotopy class of K. Let $p \in \mathbb{R}^3$ be a point not on K and let $\Lambda_p \subset ST^*\mathbb{R}^3$ denote the Legendrian conormal sphere of p. We consider certain filtered quotients of $CE(\Lambda_K \cup \Lambda_p)$, called R_{Kp}, R_{pK}, and R_{KK}, together with a product operation $m\colon R_{Kp} \otimes R_{pK} \to R_{KK}$, borrowed from wrapped Floer cohomology.

Theorem 1.1. *Ekholm, Ng, and Shende [2017a, Theorem 1.1] Two knots $K, J \subset \mathbb{R}^3$ are isotopic if and only if the triples (R_{Kp}, R_{pK}, R_{KK}) and (R_{Jp}, R_{pJ}, R_{JJ}), with the product m, are quasi-isomorphic. It follows in particular that Λ_K and Λ_J are (parameterized) Legendrian isotopic if and only if K and J are isotopic.*

A version of this theorem was first proved by Shende [2016] using micro-local sheaves and was reproved using holomorphic disks in Ekholm, Ng, and Shende [2017a]. We point out that the Legendrian conormal tori of any two knots are smoothly isotopic when considered as ordinary submanifolds of $ST^*\mathbb{R}^3$. Theorem 1.1 and its relations to string topology, Floer cohomology, and micro-local sheaves are discussed in Section 3.

1.2 Physical aspects of knot contact homology. We start from Witten's relation between Chern–Simons gauge theory and open topological string Witten [1995] together with Ooguri–Vafa's study of large N duality for conormals of knots Ooguri and Vafa [2000, 2002]. Let M be a closed 3-manifold. Witten identified the partition function of $U(N)$ Chern–Simons gauge theory on M with the partition function of open topological string on T^*M with N branes on the Lagrangian zero-section M. In Chern–Simons theory, the n-colored HOMFLY-PT polynomial of a knot $K \subset M$ equals the expectation value of the holonomy around the knot of the $U(N)$-connection in the n^{th} symmetric representation. The generating function of n-colored HOMFLY-PT polynomials correspond on the string side to the partition function of open string theory in T^*M with N branes on M and one brane on the conormal L_K of the knot.

For $M = S^3$, large N duality says that the open string in T^*S^3 with N-branes on S^3 is equivalent to the closed string, or Gromov–Witten theory, in the non-compact Calabi–Yau manifold X which is the total space of the bundle $\mathcal{O}(-1)^{\oplus 2} \to \mathbb{C}P^1$ (the resolved conifold), provided area$(\mathbb{C}P^1) = Ng_s$, where g_s is the string coupling, or genus, parameter. As smooth manifolds, $X - \mathbb{C}P^1$ and $T^*S^3 - S^3$ are diffeomorphic. As symplectic manifolds they are closely related, in particular both are asymptotic to $[0,\infty) \times ST^*S^3$ at infinity.

If $K \subset S^3$ is a knot then after a non-exact shift, see Koshkin [2007], $L_K \subset T^*S^3 - S^3$, and we can view L_K as a Lagrangian submanifold in X. This leads to the following relation between the colored HOMFLY-PT polynomial and open topological string or open Gromov–Witten theory in X. Let $C_{\chi,r,n}$ be the count of (generalized) holomorphic curves in X with boundary on L_K, of Euler characteristic χ, in relative homology class $rt + nx$, where t is the class of $[\mathbb{C}P^1] \in H_2(X, L_K)$ and $x \in H_2(X, L_K)$ maps to the generator of $H_1(L_K)$ under the connecting homomorphism. If

$$F_K(e^x, g_s, Q) = \sum_{n,r,\chi} C_{n,r,\chi}\, g_s^{-\chi} Q^r e^{nx},$$

then

$$\Psi_K(x) := e^{F_K(x)} = \sum H_{K,n}(q,Q) e^{nx}, \quad q = e^{g_s},\ Q = q^N,$$

where $H_{K,n}$ denotes the n-colored HOMFLY-PT polynomial of K.

The colored HOMFLY-PT polynomial is q-holonomic Garoufalidis, Lauda, and Le [2016], which in our language can be expressed as follows. Let $e^{\hat{x}}$ denote the operator which is multiplication by e^x and $e^{\hat{p}} = e^{g_s \frac{\partial}{\partial x}}$. Then there is a polynomial $\hat{A}_K = \hat{A}_K(e^{\hat{x}}, e^{\hat{p}})$ such that $\hat{A}_K \Psi_K = 0$.

We view Q as a parameter and think of it as fixed. Then from the short-wave asymptotic expansion of the wave function Ψ_K,

$$\Psi_K(x) = e^{F_K} = \exp\left(g_s^{-1} W_K^0(x) + W_K^1(x) + g_s^{j-1} W_K^j(x) + \dots\right),$$

we find that $p = \frac{\partial W_K^0}{\partial x}$ parameterizes the algebraic curve $\{A_K(e^x, e^p) = 0\}$, where the polynomial A_K is the classical limit $g_s \to 0$ of the operator polynomial $\hat{A}_K$. In terms of Gromov–Witten theory, $W_K(x) = W_K^0(x)$ can be interpreted as the disk potential, the count of holomorhic disks ($\chi = 1$ curves) in X with boundary on L_K.

In Aganagic and Vafa [2012] it was observed (in computed examples) that the polynomial A_K agreed with the *augmentation polynomial* Aug_K of knot contact homology. To describe that polynomial, we consider a version $\mathcal{A}_K$ of $CE(\Lambda_K)$ with coefficients in the group algebra of the second relative homology $\mathbb{C}[H_2(ST^*S^3, \Lambda_K)] \approx \mathbb{C}[e^{\pm x}, e^{\pm p}, Q^{\pm 1}]$, where x and p map to the longitude and meridian generators of $H_1(\Lambda_K)$, and $Q = e^t$ for

$t = [ST_p^* S^3]$, the class of the fiber sphere. If $\mathbb{C}$ is considered as a dg-algebra in degree 0 then the *augmentation variety* V_K is the closure of the set in the space of coefficients where there is a chain map into $\mathbb{C}$:

$$V_K = \text{closure}\big(\{(e^x, e^p, Q) \colon \text{there exists a chain map } \epsilon \colon \mathcal{A}_K \to \mathbb{C}\}\big),$$

and the *augmentation polynomial* Aug_K is its defining polynomial. We have the following result that connects knot contact homology and Gromov–Witten theory at the level of the disk.

Theorem 1.2. *Aganagic, Ekholm, Ng, and Vafa [2014, Theorem 6.6 and Remark 6.7] If $W_K(x)$ is the Gromov–Witten disk potential of $L_K \subset X$ then $p = \frac{\partial W_K}{\partial x}$ parameterizes a branch of the augmentation variety V_K.*

The augmentation polynomial Aug_K of a knot K is obtained by elimination theory from explicit polynomial equations. Theorem 1.2 thus leads to a rather effective indirect calculation of the Gromov–Witten disk potential. It is explained in Section 4.

In Section 5 we discuss the higher genus counterpart of Theorem 1.2. We sketch the construction of a higher genus generalization of knot contact homology that we call *Legendrian SFT*. In this theory, the operators $e^{\hat{x}}$ and $e^{\hat{p}}$ have natural enumerative geometrical interpretations. Furthermore, in analogy with the calculation of the augmentation polynomial, elimination theory in the non-commutative setting should give the operator polynomial $\widehat{\text{Aug}}_K(e^{\hat{x}}, e^{\hat{p}})$ such that $\widehat{\text{Aug}}_K \Psi_K = 0$, and thus determine the recursion relation for the colored HOMFLY-PT.

Remark 1.3. Theorem 1.2 and other results about open Gromov–Witten theory presented here should be considered established from the physics point of view. From a more strict mathematical perspective, they are not rigorously proved and should be considered as conjectures.

Acknowledgments. I am much indebted to my coauthors, Aganagic, Cieliebak, Etnyre, Latchev, Lekili, Ng, Shende, Sullivan, and Vafa, of the papers on which this note is based.

2 Knot contact homology and Chekanov–Eliashberg dg-algebras

In this section we introduce Chekanov–Eliashberg dg-algebras in the cases we use them.

2.1 Background notions. Let M be an orientable 3-manifold and consider the unit cotangent bundle ST^*M with the contact 1-form α which is the restriction of the action form pdq. The hyperplane field $\xi = \ker(\alpha)$ is the contact structure determined by α and $d\alpha$ gives a symplectic form on ξ. The first Chern-class of ξ vanishes, $c_1(\xi) = 0$.

Let $\Lambda \subset ST^*M$ be a Legendrian submanifold, $\alpha|_\Lambda = 0$. Then the tangent spaces of Λ are Lagrangian subspaces of ξ. Since $c_1(\xi) = 0$ there is a Maslov class in $H^1(\Lambda; \mathbb{Z})$ that measures the total rotation of $T\Lambda$ in ξ. Here we will consider only Legendrian submanifolds with vanishing Maslov class.

The *Reeb vector field* R of α is characterized by $d\alpha(\cdot, R) = 0$ and $\alpha(R) = 1$. Flow segments of R that begin and end on Λ are called *Reeb chords*. The Reeb flow on ST^*M is the lift of the geodesic flow on M. Consequently, if $K \subset M$ is a knot (or any submanifold) then Reeb chords of Λ_K correspond to geodesics connecting K to itself and perpendicular to K at its endpoints.

2.2 Coefficients in chains on the based loop space. Let $M = \mathbb{R}^3$, $K \subset \mathbb{R}^3$ be a knot and $p \in \mathbb{R}^3 - K$ a point. Let $\Lambda_0 = \Lambda_p$, $\Lambda_1 = \Lambda_K$, and $\Lambda = \Lambda_0 \cup \Lambda_1$. The algebra $CE(\Lambda)$ is generated by the Reeb chords of Λ and homotopy classes of loops in Λ. We define the coefficient ring $\mathbf{k}_\Lambda$ as the algebra over $\mathbb{C}$ generated by idempotents e_j corresponding to Λ_j so that $e_i e_j = \delta_{ij} e_i$, $i, j \in \{0, 1\}$, where δ_{ij} is the Kronecker delta.

Note that Λ_0 is a sphere and Λ_1 is a torus. Fix generators λ and μ of $\pi_1(\Lambda_1)$ (corresponding to the longitude and the meridian of K) and think of them as generators of the group algebra $\mathbb{C}[\pi_1(\Lambda_1)] \approx \mathbb{C}[\lambda^{\pm 1}, \mu^{\pm 1}]$. We let $CE(\Lambda)$ be the algebra over $\mathbf{k}_\Lambda$ generated by Reeb chords c, and the homotopy classes λ and μ. The generators λ and μ satisfy the relations in the group algebra and the following additional relations hold:

$$ce_j = \begin{cases} c & \text{if } c \text{ starts on } \Lambda_j, \\ 0 & \text{otherwise,} \end{cases} \qquad e_k c = \begin{cases} c & \text{if } c \text{ ends on } \Lambda_k, \\ 0 & \text{otherwise,} \end{cases}$$
$$e_j \lambda_k = \lambda_k e_j = \delta_{jk} \lambda_k, \qquad e_j \mu_k = \mu_k e_j = \delta_{jk} \mu_k.$$

The grading of λ and μ is $|\lambda| = |\mu| = 0$ and Reeb chords are graded by the Conley–Zehnder index, which in the case of knot contact homology equals the Morse index of the underlying binormal geodesic, see Ekholm, J. B. Etnyre, Ng, and M. G. Sullivan [2013]. We can thus think of elements of $CE(\Lambda)$ as finite linear combinations of composable monomials $\mathbf{c}$ of the form

$$\mathbf{c} = \gamma_0 c_1 \gamma_1 c_2 \gamma_2 \dots \gamma_{m-1} c_m \gamma_m,$$

where γ_j is a homotopy class of loops in Λ and c_{j+1} is a Reeb chord, and composable means that c starts at the component of γ_j and ends at the component of γ_{j-1}. We then have the decomposition

$$CE(\Lambda) = \bigoplus_{i,j} CE(\Lambda)_{i,j},$$

where $CE(\Lambda)_{i,j}$ is generated by monomials which start on Λ_j and ends on Λ_i. The product of two monomials is given by concatenation if the result is composable and zero otherwise.

The differential is defined to be 0 on e_i and on elements of $\mathbb{Z}[\pi_1(\Lambda_1)]$ and is given by a holomorphic disk count on Reeb chord generators that we describe next. Fix a complex structure J on the symplectization $\mathbb{R} \times ST^*\mathbb{R}^3$, with symplectic form $d(e^t\alpha)$, $t \in \mathbb{R}$, that is invariant under the $\mathbb{R}$-translation and maps ξ to itself. If c is a Reeb chord then $\mathbb{R} \times c$ is a holomorphic strip with boundary on the Lagrangian submanifold $\mathbb{R} \times \Lambda$. Fix a base point in each component of Λ and fix for each Reeb chord endpoint a reference path connecting it to the base point. Consider a Reeb chord a and a composable word $\mathbf{b}$ of homotopy classes and Reeb chords of the form

$$\mathbf{b} = \gamma_0 b_1 \gamma_1 b_2 \gamma_2 \dots \gamma_{m-1} b_m \gamma_m,$$

where γ_0 lies in the component where a ends and γ_m in the component where a starts. We let $\mathfrak{M}(a;\mathbf{b})$ denote the moduli space of holomorphic disks

$$u\colon (D, \partial D) \to (\mathbb{R} \times ST^*\mathbb{R}^3, \mathbb{R} \times \Lambda), \quad du + J \circ du \circ i = 0,$$

with one positive and m negative boundary punctures, which are asymptotic to the Reeb chord strip $\mathbb{R} \times a$ at positive infinity at the positive puncture and to the Reeb chord strip $\mathbb{R} \times b_j$ at negative infinty at the j^{th} negative puncture and such that the closed off path between punctures j and $j+1$ lies in homotopy class γ_j, where puncture 0 and $m+1$ both refer to the positive puncture, see Figure 1. The dimension of the moduli space $\mathfrak{M}(a;\mathbf{b})$ equals $|a| - |\mathbf{b}|$.

We define

$$\partial a = \sum_{a - |\mathbf{b}| = 1} |\mathfrak{M}(a;\mathbf{b})| \mathbf{b}, \tag{1}$$

where $|\mathfrak{M}(a;\mathbf{b})|$ denotes the algebraic number of $\mathbb{R}$-families of disks in $\mathfrak{M}(a;\mathbf{b})$ and extend to monomials by Leibniz rule. For the count in (1) to make sense we need the solutions to be transversely cut out. Since disks with one positive puncture cannot be multiple covers, transversality is relatively straightforward. Furthermore, the sum is finite by the SFT version of Gromov compactness.

The basic result for Chekanov–Elisahberg algebras is then the following.

Lemma 2.1. *The map ∂ is a differential, $\partial \circ \partial = 0$ and the quasi-isomorphism class of $CE(\Lambda)$ is invariant under Legnedrian isotopies of Λ. Furthermore, the differential respects the decomposition $CE(\Lambda) = \bigoplus_{i,j} CE(\Lambda)_{i,j}$ which thus descends to homology.*

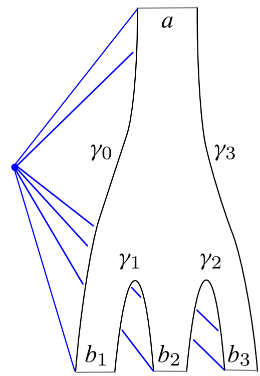


Figure 1: A disk contributing $\gamma_0 b_1 \gamma_1 b_2 \gamma_2 b_3 \gamma_3$ to ∂a.

Remark 2.2. For general contact manifolds, $CE(\Lambda)$ is an algebra over the so called orbit contact homology algebra. In the cases under study, $ST^*\mathbb{R}^3$ and ST^*S^3, the orbit contact homology algebra is trivial in degree 0 and can be neglected.

Remark 2.3. For general Legendrian submanifolds Λ, the version of $CE(\Lambda)$ considered here is more complicated. The group ring generators for torus components are replaced by chains on the based loop space of the corresponding components and moduli spaces of all dimensions contribute to the differential, see Ekholm and Lekili [2017].

Sketch of proof. If a is a Reeb chord then $\partial(\partial a)$ counts two level curves joined at Reeb chords. By gluing and SFT compactness such configurations constitute the boundary of an oriented 1-manifold and hence cancel algebraically. The invariance property can be proved in a similar way by looking at the boundary of the moduli space of holomorphic disks in Lagrangian cobordisms associated to Legendrian isotopies. See e.g. Ekholm, J. Etnyre, and M. Sullivan [2007] for details. □

2.3 Coefficients in relative homology. Our second version of the Chekanov–Eliashberg dg-algebra of the conormal $\Lambda_K \subset ST^*S^3$ of a knot $K \subset S^3$ is denoted $\mathfrak{A}_K$. The algebra $\mathfrak{A}_K$ is generated by Reeb chords graded as before. Its coefficient ring is the group algebra $\mathbb{C}[H_2(ST^*S^3, \Lambda_K)]$ and group algebra elements commute with Reeb chords. To define the differential we fix for each Reeb chord a disk filling the reference paths. Capping off punctured disks in the moduli space $\mathfrak{M}(a, \mathbf{b})$ with these disks we get a relative homology class and define the differential on Reeb chord generators of $\mathfrak{A}_K$ as

$$da = \sum_{|a|-|\mathbf{c}|=1} |\mathfrak{M}(a; \mathbf{c})|\mathbf{c}.$$

Here $\mathbf{c} = e^A c_1 \dots c_m$, where c_j are the Reeb chords at the negative punctures of the disks in the moduli space and $A \in H_2(ST^*S^3; \Lambda_K)$ is the relative homology class of the capped off disks. That d is a differential and the quasi-isomorphism invariance of $\mathcal{Q}_K$ under Legendrian isotopies follow as before.

2.4 Knot contact homology in basic examples. We calculate the knot contact homology dg-algebras (in the lowest degrees) for the unknot and the trefoil knot. For general formulas we refer to Ekholm, J. B. Etnyre, Ng, and M. G. Sullivan [2013] and Ekholm, J. Etnyre, Ng, and M. Sullivan [2013]. The expressions give the differential in $\mathcal{Q}_K$. for the differential in $CE(\Lambda_K)$, set $Q = 1$, $e^x = \lambda$, and $e^p = \mu$.

2.4.1 The unknot. Representing the unkot as a round circle in the plane we find that it has an S^1-Bott family of binormal geodesics and correspondingly an S^1-Bott family of Reeb chords. After small perturbation this gives two Reeb chords c and e of degrees $|c| = 1$ and $|e| = 2$. The differential can be computed using Morse flow trees, see Ekholm [2007] and Ekholm, J. B. Etnyre, Ng, and M. G. Sullivan [2013]. The result is

$$de = 0, \quad dc = 1 - e^x - e^p - Qe^x e^p. \tag{2}$$

2.4.2 The trefoil knot. Represent the trefoil knot as a 2-strand braid around the unkot. If the trefoil T lies sufficiently close to the unkot U, then its conormal torus Λ_T lies in a small neighborhood $N(\Lambda_U)$ of the unknot conormal, which can be identified with the neighborhood the zero section in its 1-jet space $J^1(\Lambda_U)$. The projection $\Lambda_T \to \Lambda_U$ is a 2-fold cover and holomorphic disks with boundary on $\mathbb{R} \times \Lambda_T$ correspond to holomorphic disks on Λ_U with flow trees attached, where the flow trees are determined by $\Lambda_T \subset J^1(\Lambda_U)$, see Ekholm, J. B. Etnyre, Ng, and M. G. Sullivan [2013]. This leads to the following description of $\mathcal{Q}_T$ in degrees ≤ 1. The Reeb chords are:

degree 1: $b_{12}, b_{21}, c_{11}, c_{12}, c_{21}, c_{22}$, degree 0: a_{12}, a_{21},

with differentials

$$\begin{aligned}
dc_{11} &= e^x e^p - e^x - (2Q - e^p)a_{12} - Qa_{12}^2 a_{21}, & dc_{12} &= Q - e^p + e^p a_{12} + Qa_{12}a_{21},\\
dc_{21} &= Q - e^p - e^x e^p a_{21} + Qa_{12}a_{21}, & dc_{22} &= e^p - 1 - Qa_{21} + e^p a_{12}a_{21},\\
db_{12} &= e^{-x} a_{12} - a_{21}, & db_{21} &= a_{21} - e^x a_{12}.
\end{aligned}$$

3 A complete knot invariant

In this section we discuss the completeness of knot contact homology as a knot invariant and describe its relations of to string topology, wrapped Floer cohomology, and microlocal sheaves.

3.1 Filtered quotients and a product. We use notation as in Section 2.2, $\Lambda = \Lambda_p \cup \Lambda_K = \Lambda_0 \cup \Lambda_1$, and consider $CE(\Lambda)$. The group ring $\mathbb{Z}[\pi_1(\Lambda_K)]$ is a subalgebra of $CE(\Lambda)$ generated by the longitude and meridian generators $\lambda^{\pm 1}$ and $\mu^{\pm 1}$. Other generators are Reeb chords that correspond to binormal geodesics. If γ is a geodesic we write c for the corresponding Reeb chord. The grading of Reeb chords with endpoints on the same connected component is well-defined, while the grading for *mixed chords* connecting distinct components are defined only up to an over all shift specified by a cetrain reference path connecting the two components. Let $\operatorname{ind}(\gamma)$ denote the Morse index of the geodesic γ.

Lemma 3.1. *Ekholm, Ng, and Shende [2017a, Proposition 2.3] There is a choice of reference path so that the grading in $CE(\Lambda)$ of a Reeb chord c corresponding to the geodesic γ is as follows: if c connects Λ_K to Λ_K or Λ_p to Λ_K then $|c| = \operatorname{ind}(\gamma)$, and if c connects Λ_K to Λ_p then $|c| = \operatorname{ind}(\gamma) + 1$.*

Consider the filtration on $CE(\Lambda)$ by the number of mixed Reeb chords, and the corresponding filtered quotients:

$$CE(\Lambda)_{1,1} = \mathfrak{F}^0_{11} \supset \mathfrak{F}^2_{11} \supset \mathfrak{F}^4_{11} \supset \dots, \quad CE^{(2k)}_{11} = \mathfrak{F}^{2k}_{11} / \mathfrak{F}^{2k+2}_{11},$$
$$CE(\Lambda)_{i,j} = \mathfrak{F}^1_{ij} \supset \mathfrak{F}^3_{ij} \supset \mathfrak{F}^5_{ij} \supset \dots, \quad CE^{(2k+1)}_{ij} = \mathfrak{F}^{2k+1}_{ij} / \mathfrak{F}^{2k+3}_{ij}, \text{ for } i \neq j,$$

where $\mathfrak{F}^r$ denotes the subalgebra generated by monomials with at least r mixed Reeb chords. The differential respects this filtration. Lemma 3.1 shows that $CE(\Lambda)$ is supported in non-negative degrees and that monomials of lowest degree $d(i,j) \in \{0,1\}$ in $CE(\Lambda)_{i,j}$ contain the minimal possible number $s(i,j) \in \{0,1\}$ of mixed Reeb chords. We then find that $H_{d(i,j)}(CE(\Lambda)_{i,j}) = H_{d(i,j)}(CE^{(s(i,j))}_{ij})$. We call

$$(R_{Kp}, R_{pK}, R_{KK}) := (H_0(CE(\Lambda)_{10}), H_1(CE(\Lambda)_{01}), H_0(CE(\Lambda)_{1,1}))$$

the *knot contact homology triple* of K. The concatenation product in $CE(\Lambda)$ turns R_{Kp} and R_{pK} into left and right modules, respectively, and R_{KK} into a left-right module over $\mathbb{Z}[\lambda^{\pm 1}, \mu^{\pm 1}]$.

We next consider a product for the knot contact homology triple that is closely related to the product in wrapped Floer cohomology. As the differential, it is defined in terms of moduli spaces of holomorphic disk but for the product there are two positive punctures rather than one.

Let a and b be Reeb chords connecting Λ_p to Λ_K and vice versa. Let $\mathbf{c}$ be a monomial in $CE(\Lambda_K)$. Define $\mathfrak{M}(a,b;\mathbf{c})$ as the moduli space of holomorphic disks $u\colon D \to \mathbb{R}\times T^*\mathbb{R}^3$ with two positive punctures asymptotic to a and b, such that the boundary arc between them maps to $\mathbb{R}\times\Lambda_p$, and such that the remaining punctured arc in the boundary maps to Λ_K with homotopy class and negative punctures according to $\mathbf{c}$. We then have

$$\dim(\mathfrak{M}(a,b;\mathbf{c})) = |a|+|b|-|\mathbf{c}|.$$

Define

$$m'(a,b) = \sum_{|\mathbf{c}|=|a|+|b|-1} |\mathfrak{M}(a,b;\mathbf{c})|\mathbf{c}$$

and use this to define the chain level product $m\colon CE(\Lambda)^{(1)}_{10}\otimes CE(\Lambda)^{(1)}_{01} \to CE(\Lambda)^{(0)}_{11}$ as $m(\mathbf{a}a, b\mathbf{b}) = \mathbf{a}m'(a,b)\mathbf{b}$.

Proposition 3.2. *Ekholm, Ng, and Shende [2017a, Proposition 2.13] The product m descends to homology and gives a product $m\colon R_{Kp}\otimes R_{pK} \to R_{KK}$. The knot contact homology triple as modules over $\mathbb{Z}[\pi_1(\Lambda_K)]$ and with the product m is invariant under Legendrian isotopy.*

3.2 String topology and the cord algebra. In this section we define a topological model for knot contact homology in low degrees that one can think of as the string topology of a certain singular space. Our treatment will be brief and we refer to Cieliebak, Ekholm, Latschev, and Ng [2017] and Ekholm, Ng, and Shende [2017a] for full details.

Let $K\subset\mathbb{R}^3$ be a knot and $p\in\mathbb{R}^3-K$ a point with Lagrangian conormals L_K and L_p. Let Σ be the union $\Sigma = \mathbb{R}^3\cup L_K\cup L_p \subset T^*\mathbb{R}^3$. Pick an almost complex structure J compatible with the metric along the zero section. Fix base points $x_K\in L_K-\mathbb{R}^3$ and $x_p\in L_p-\mathbb{R}^3$.

We consider broken strings which are paths $s\colon[a,b]\to\Sigma$ that connect base points, $c(a), c(b)\in\{x_p,x_K\}$ and that admit a subdivision $a<t_1<\dots<t_m<b$ such that $s|_{[t_i,t_{i+1}]}$ is a C^k-map into one of the irreducible components of Σ and such that the left and right derivatives at switches (i.e., points where c switches irreducible components) are related by $\dot c(t_j-) = J\dot c(t_j+)$.

For $\ell>0$, let Σ_ℓ denote the space of strings with ℓ switches at p and with the C^k-topology for some $k>0$. Write $\Sigma_\ell = \Sigma_\ell^{KK}\cup\Sigma_\ell^{Kp}\cup\Sigma_\ell^{pK}\cup\Sigma_\ell^{pp}$, where Σ_ℓ^{KK} denotes strings that start and end at x_K, etc. For $d>0$, let

$$C_d(\Sigma_\ell) = C_d(\Sigma_\ell^{KK})\oplus C_d(\Sigma_\ell^{Kp})\oplus C_d(\Sigma_\ell^{pK})\oplus C_d(\Sigma_\ell^{pp})$$

denote singular d-chains of Σ_ℓ in general position with respect to K. We introduce two string topology operations associated to K, $\delta_K^Q,\delta_K^N\colon C_k(\Sigma_\ell)\to C_{k-1}(\Sigma_{\ell+1})$. If σ is a generic d-simplex then $\delta_K^Q(\sigma)$ is the chain parameterized by the locus in σ of strings with components in S^3 that intersect K at interior points. The operation splits the curve at such intersection and inserts a spike in L_K, see Cieliebak, Ekholm, Latschev, and Ng [2017]. The operation δ_K^Q is defined similarly exchanging the role of $\mathbb{R}^3$ and L_K. There are also similar operations $\delta_p^Q, \delta_p^N : C_k(\Sigma_\ell) \to C_{k-2}(\Sigma_{\ell+1})$ at p that will play less of a role here.

Let ∂ denote the singular differential on $C_*(\Sigma_\ell)$ and let $C_m = \bigoplus_{k+\ell/2=m} C_k(\Sigma_\ell)$. We introduce a Pontryagin product which concatenates strings at p. We write R_{KK}^{st}, R_{Kp}^{st}, and R_{pK}^{st} for the degree 0 homology of the corresponding summands of C_*.

Proposition 3.3. *Cieliebak, Ekholm, Latschev, and Ng [2017] and Ekholm, Ng, and Shende [2017a] The map $d = \partial + \delta_K^Q + \delta_K^N + \delta_p^Q + \delta_p^N$ is a differential on C_*. The homology of d in degree 0 is the cokernel of $\partial + \delta_K^Q + \delta_K^N : C_1 \to C_0$ (where δ_p^Q and δ_p^N vanishes for degree reasons) and is as follows:*

$$R_{KK}^{\mathrm{st}} \approx \hat{R} + R(1-\mu), \quad R_{Kp}^{\mathrm{st}} \approx R, \quad R_{pK}^{\mathrm{st}} \approx R(1-\mu),$$

where $R = \mathbb{Z}[\pi_1(\mathbb{R}^3 - K)]$ and $\hat{R} = \mathbb{Z}[\pi_1(\Lambda_K)]$.

We next consider a geometric chain map of algebras $\Phi\colon CE(\Lambda) \to C_*$, where the multiplication on C_* is given by chain level concatenation of broken strings. The map is defined as follows on generators. If a is a Reeb chord let $\mathfrak{M}(a;\Sigma)$ denote the moduli space of holomorphic disks in T^*S^3 with boundary on Σ and Lagrangian intersection punctures at K. The evaluation map gives a chain of broken strings for each $u \in \mathfrak{M}(a;\Sigma)$. Let $[\mathfrak{M}(a;\Sigma)]$ denote the chain of broken strings carried by the moduli space and define $\Phi(a) = [\mathfrak{M}(a;\Sigma)]$.

Proposition 3.4. *Ekholm, Ng, and Shende [2017a] The map Φ is a chain map. It induces an isomorphism*

$$(R_{Kp}, R_{pK}, R_{KK}) \to \left(R_{Kp}^{\mathrm{st}}, R_{pK}^{\mathrm{st}}, R_{KK}^{\mathrm{st}}\right)$$

that intertwines the product m and the Pontryagin product at p.

Proof of Theorem 1.1. Propositions 3.3 and 3.4 imply that the knot contact homology triple knows the group ring of the knot group and the action of λ and μ. Properties of left-orderable groups together with Waldhausen's theorem then give the result, see Ekholm, Ng, and Shende [ibid.] for details. □

3.3 Partially wrapped Floer cohomology and Legendrian surgery. The knot contact homology of the previous section can also be interpreted, via Legendrian surgery, in terms of partially wrapped Floer cohomology that in turn is connected to the micro-local sheaves used by Shende [2016] to prove the completness result in Theorem 1.1. We give a very brief discussion and refer to Ekholm, Ng, and Shende [2017b, Section 6] for more details.

To a knot $K \subset \mathbb{R}^3$ we associate a Liouville sector W_K with Lagrangian skeleton $L = \mathbb{R}^3 \cup L_K$, this roughly means that L is a Lagrangian subvariety and that W_K is a regular neighborhood of L, see Sylvan [2016] and Ganatra, Pardon, and Shende [2017]. More precisely, W_K is obtained by attaching the cotangent bundle $T^*[0,\infty) \times \Lambda_K$ to $T^*\mathbb{R}^3$ along $\Lambda_K \subset ST^*\mathbb{R}^3$. We let C_K denote the cotangent fiber at $q \in [0,\infty) \times \Lambda_K$ and C_p the cotangent fiber at $p \in \mathbb{R}^3$. Such handle attachments were considered in Ekholm and Lekili [2017] where it was shown that there exists a natural surgery quasi-isomorphism $\Phi \colon CE(\Lambda_K) \to CW^*(C_K)$, where CW^* denotes wrapped Floer cohomology. There are directly analogous quasi-isomorphisms

$$CE_{01}^{(1)} \to CW^*(C_K, C_p), \quad CE_{10}^{(1)} \to CW^*(C_p, C_K), \quad CE_{11}^{(0)} \to CW^*(C_p, C_p),$$

under which the product m corresponds to the usual triangle product m_2 on CW^*.

In Shende [2016], the conormal torus Λ_K of a knot $K \subset \mathbb{R}^3$ was studied via the category of sheaves microsupported in L. This sheaf category can also be described as the category of modules over the wrapped Fukaya category of W_K which is generated by the two cotangent fibers C_K and C_p. The knot contact homology triple with m then have a natural interpretation as calculating morphisms in a category equivalent to that studied in Shende [ibid.].

4 Augmentations and the Gromov–Witten disk potential

Let $K \subset S^3$ be a knot and let L_K denote its conormal Lagrangian. Shifting L_K along the 1-form dual to its unit tangent vector we get a non-exact Lagrangian that is disjoint from the 0-section. We identify the complement of the 0-section in T^*S^3 with the complement of the 0-section in the resolved conifold X. Under this identification, L_K becomes a uniformly tame Lagrangian, see Koshkin [2007], which is asymptotic to $\mathbb{R} \times \Lambda_K \subset \mathbb{R} \times ST^*S^3$ at infinity. The first condition implies that L_K can be used as boundary condition for holomorphic curves and the second that at infinity, holomorphic curves on (X, L_K) can be identified with the $\mathbb{R}$-invariant holmorphic curves of $(\mathbb{R} \times ST^*S^3, \mathbb{R} \times \Lambda_K)$.

Since $c_1(X) = 0$ and the Maslov class of L_K vanishes, the formal dimension of any holomorphic curve in X with boundary on L_K equals 0. Fixing a perturbation scheme one then gets a 0-dimensional moduli space of curves. Naively, the open Gromov–Witten invariant of L_K would be the count of these rigid curves. Simple examples however show that such a count is not invariant under deformations, contradicting what topological string theory predicts.

To resolve this problem on the Gromov–Witten side, we count more involved configurations of curves that we call *generalized curves*. In this section we consider the simpler case of disks and then in Section 5 the case of general holomorphic curves. The problems of open Gromov–Witten theory in this setting was studied from the mathematical perspective also by Iacovino [2009a,b]. From the physical perspective, the appearance of more complicated configurations then bare holomorphic curves seems related to boundary terms in the path integral localized on the moduli space of holomorphic curves with boundary which, unlike in the case of closed curves, has essential codimension one boundary strata.

4.1 Augmentations of non-exact Lagrangians and disk potentials. We will construct augmentations induced by the non-exact Lagrangian filling $L_K \subset X$. In order to explain how this works we first consider the case of the exact filling $L_K \subset T^*S^3$. The exact case is a standard ingredient in the study of Chekanov–Eliashberg dg-algebras, see e.g. Ekholm, Honda, and Kálmán [2016]. Consider the algebra $\mathcal{A}_K$ with coefficients in $\mathbb{C}[e^{\pm x}, e^{\pm p}, Q^{\pm 1}]$. Here we set $e^p = 1$ since p bounds in L_K and $Q = 1$ since the cotangent fiber sphere bounds in ST^*S^3. If a is a Reeb chord of Λ_K, we let $\mathcal{M}_n(a)$ denote the moduli space of holomorphic disks with positive puncture at a and boundary on L_K that lies in the homology class nx. Then $\dim(\mathcal{M}_n(a)) = |a|$ and we define the map $\epsilon_0 \colon \mathcal{A}_K \to \mathbb{C}[e^{\pm x}]$ on degree 0 Reeb chords a as

$$\epsilon_0(a) = \sum_n |\mathcal{M}_n(a)| e^{nx}.$$

Lemma 4.1. *The map $\epsilon_0 \colon \mathcal{A}_K|_{Q=1, e^p=1} \to \mathbb{C}[e^{\pm x}]$ is a chain map, $\epsilon_0 \circ d = 0$.*

Proof. Configurations contributing to $\epsilon_0 \circ d$ are two level broken disks that are in one to one correspondence with the boundary of the oriented 1-manifolds $\mathcal{M}_n(c)$, $|c| = 1$. □

We next consider the case of the non-exact Lagrangian filling $L_K \subset X$. In this case, $Q = e^t$, where $t = [\mathbb{C}P^1] \in H_2(X)$ and we look for a chain map $\mathcal{A}_K \to \mathbb{C}[e^{\pm x}, Q^{\pm 1}]$. If a is a Reeb chord, then let $\mathcal{M}_{r,n}(a)$ denote the moduli space of holomorphic disks in X with boundary on L_K in relative homology class $rt + nx$.

Consider first the naive generalization of the exact case and define

$$\epsilon'(a) = \sum_{r,n} |\mathcal{M}_{r,n}(a)| Q^r e^{nx}.$$

We look at the boundary of 1-dimensional moduli spaces $\mathcal{M}_{r,n}(c)$, $|c| = 1$. Unlike in the exact case, two level broken curves do not account for the whole boundary of $\mathcal{M}_{r,n}(c)$ and consequently the chain map equation does not hold. The reason is that there are non-constant holomorphic disks without positive punctures on L_K and a 1-dimensional family of disks can split off non-trivial such disks under so called boundary bubbling. Together with two level disks, disks with boundary bubbles account for the whole boundary of the moduli spacce.

The problem of boundary bubbling is well-known in Floer cohomology and was dealt with there using the method of bounding cochains introduced by Fukaya, Oh, Ohta, Fukaya, Oh, Ohta, and Ono [2009]. We implement this method in the current set up by introducing non-compact bounding chains (with boundary at infinity) as follows. We use a perturbation scheme to make rigid disks transversely cut out energy level by energy level. For each transverse disk u we also fix a bounding chain σ_u, i.e., σ_u is a non-compact 2-chain in L_K that interpolates between the boundary ∂u and a multiple of a fixed curve ξ in Λ_K in the longitude homology class $x \in H_1(\Lambda_K)$ at infinity. This allows us to define the Gromov–Witten disk potential as a sum over finite trees Γ, where there is a rigid disk u_v at each vertex $v \in \Gamma$ and for every edge connecting vertices v and v' there is an intersection point between ∂u_v and $\sigma_{v'}$ weighted by $\pm\frac{1}{2}$, according to the intersection number. We call such a tree a *generalized disk* and define the Gromov–Witten disk potential $W_K(x, Q)$ as the generating function of generalized disks.

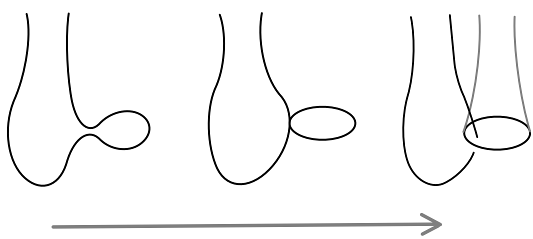


Figure 2: Bounding chains turn boundary breaking into interior points in moduli spaces: the disk family continues as a family of disks with the bounding chain inserted.

We then define $\mathfrak{M}'_{r,n}(a)$ as the moduli space of holomorphic disks with positive puncture at a and with insertion of bounding chains of generalized disks along its boundary such that the total homology class of the union of all disks in the configuration lies in the class $rt + nx$. Let $\epsilon\colon \mathfrak{A}_K \to \mathbb{C}[e^{\pm x}, Q^{\pm 1}]$ be the map

$$\epsilon(a) = \sum_{r,n} |\mathfrak{M}'_{r,n}(a)| Q^r e^{nx}.$$

Proposition 4.2. *If*

$$p = \frac{\partial W_K}{\partial x} \tag{3}$$

then ϵ is a chain map, $\epsilon \circ d = 0$. Consequently, (3) parameterizes a branch of the augmentation variety and Theorem 1.2 follows.

Proof. The bounding chains are used to remove boundary bubbling from the boundary of the moduli space, see Figure 2. As a consequence the boundary of the moduli space of disks with one positive puncture and with insertions of generalized disks correspond to two level disks with insertions. It remains to count the disks at infinity with insertions. At infinity all bounding chains are multiples of the longitude generator x. A bounding chain going n-times around x can be inserted nm times in a curve that goes m times around p. It follows that the substitution $e^p = e^{\frac{\partial W_K}{\partial x}}$ corresponds to counting disks with insertions. □

Corollary 4.3. *The Gromov–Witten disk potential W_K is an analytic function.*

Proof. The defining equation of the augmentation variety can be found from the knot contact homology differential by elimination theory. It is therefore an algebraic variety and the Gromov–Witten disk potential in (3) is an analytic function. □

4.2 Augmentation varieties in basic examples.

We calculate the augmentation variety from the formulas in Section 2.4.

4.2.1 The unknot.

The augmentation polynomial for the unknot U is determined directly by (2): the algebra admits an augmentation exactly when $dc = 0$ and

$$\mathrm{Aug}_U = 1 - e^x - e^p + Qe^xe^p.$$

4.2.2 The trefoil.

We need to find the locus where the right hand sides in Section 2.4.2 has common roots. The augmentation polynomial is found as:

$$\begin{aligned}
\mathrm{Aug}_T =&(e^xe^{2p} + Q^2)a_{12}(e^p(dc_{21}) - Q(dc_{22}))\\
&- (e^xe^{2p} + Q^2)(Q(dc_{21}) + e^xe^pd(c_{22}))\\
&+ e^x(e^{2p} - Q)(e^p(dc_{21}) - Q(dc_{22}))\\
&+ e^x(e^{2p} - Q)(e^xe^{2p} + Q^2)(db_{12})\\
&= e^{2x}(e^{4p} - e^{3p}) + e^x(e^{4p} - e^{3p}Q + 2e^{2p}(Q^2 - Q) - e^pQ^2 + Q^2)\\
&- (e^pQ^3 - Q^4).
\end{aligned}$$

5 Legendrian SFT and open Gromov–Witten theory

This section concerns the higher genus counterpart of the results in Section 4.

5.1 Additional geometric data for Legendrian SFT of knot conormals. We outline a definition of relevant parts of Legendrian SFT (including the open Gromov–Witten potential), for the Lagrangian conormal L_K of a knot $K \subset S^3$ in the resolved conifold, $L_K \subset X$. As in the case of holomorphic disks, see Section 4, the main point of the construction is to overcome boundary bubbling. In the disk case there is a 1-dimensional disk that interacts through boundary splitting/crossing with rigid disks. Since the moving disk is distinguished from the rigid disks, it is sufficient to use bounding chains for the rigid disks only.

In the case of higher genus curves there is no such separation. A 1-dimensional curve can boundary split on its own. To deal with this, we introduce additional geometric data that defines what might be thought of as dynamical bounding chains. We give a brief description here and refer to Ekholm and Ng [n.d.] for more details. The construction was inspired by self linking of real algebraic links (Viro's encomplexed writhe, Viro [2001]) as described in Ekholm [2002].

5.1.1 An auxiliary Morse function. Consider a Morse function $f : L_K \to \mathbb{R}$ without maximum and with the following properties. The critical points of f lie on K and are: a minimum κ_0 and an index 1 critical point κ_1. Flow lines of ∇f connecting κ_0 to κ_1 lie in K and outside a small neighborhood of K, ∇f is the radial vector field along the fiber disks in $L_K \approx K \times \mathbb{R}^2$. Note that the unstable manifold $W^{\mathrm{u}}(\kappa_1)$ of κ_1 is a disk that intersects Λ_K in the meridian cycle p.

5.1.2 A 4-chain with boundary twice L_K. Start with a 3-chain $\Gamma_K \subset ST^*S^3$ with the following properties: $\partial\Gamma_K = 2 \cdot \Lambda_K$, near the boundary Γ_K agrees with the union of small length $\epsilon > 0$ flow lines of $\pm R$ starting on Λ_K, and $\Gamma_K - \partial\Gamma_K$ is disjoint from Λ_K, see Ekholm and Ng [n.d.]. Identify $\big([0,\infty) \times ST^*S^3, [0,\infty) \times \Lambda_K\big)$ with $(X, L_K) - (\bar{X}, \bar{L}_K)$, where $(\bar{X}, \bar{L}_K)$ is compact, and let $C_K^{\infty} = [0,\infty) \times \Gamma_K$.

Consider the vector field $v(q) = \frac{\nabla f(q)}{|\nabla f(q)|}$, $q \in L_K - \{\kappa_0, \kappa_1\}$ and let G be the closure of the length $\epsilon > 0$ half rays of $\pm J v(q)$ starting at $q \in L_K$ in $\bar{L}_K$ and G' its boundary component that does not intersect L_K. A straightforward homology calculation shows that there exists a 4-chain C_K^0 in $X - L_K$ with boundary $\partial C_K^0 = G' \cup \partial C_K^{\infty}$. Define $C_K = C_K^{\infty} \times [0,\infty) \cup C_K^0 \cup G$. Then C_K is a 4-chain with regular boundary along $2 \cdot L_K$ and inward normal $\pm J\nabla f$. Furthermore, C_K intersects L_K only along its boundary and is otherwise disjoint from it. We remark that in order to achieve necessary transversality, we also need to perturb the Morse function and the chain slightly near the Reeb chord endpoints in order to avoid intersections with trivial strips, see Ekholm and Ng [ibid.].

5.2 Bounding chains for holomorphic curves. We next associate a bounding chain to each holomorphic curve $u\colon (\Sigma, \partial\Sigma) \to (X, L_K)$ in general position with respect to ∇f and C_K. Consider first the case without punctures. The boundary $u(\partial\Sigma)$ is a collection of closed curves contained in a compact subset of L_K. By general position, $u(\partial\Sigma)$ does not intersect the stable manifold of κ_1. Define σ_u' as the union of all flow lines of ∇f that starts on $u(\partial\Sigma)$. Since f has no index 2 critical points and since ∇f is vertical outside a compact, $\sigma_u' \cap (\{T\} \times \Lambda_K)$ is a closed curve, independent of T for all sufficiently large $T > 0$. Let $\partial_\infty \sigma_u' \subset \Lambda_K$ denote this curve and assume that its homology class is $nx + mp \in H_1(\Lambda_K)$. Define the bounding chain σ_u of u as

$$\sigma_u = \sigma_u' - m \cdot W^{\mathrm{u}}(\kappa_1). \tag{4}$$

Then σ_u has boundary $\partial\sigma_u = \partial u$ and boundary at infinity $\partial^\infty \sigma_u$ in the class $nx + 0p$.

Consider next the general case when $u\colon (\Sigma, \partial\Sigma) \to (X, L_K)$ has punctures at Reeb chords $c_1, \dots, c_m$. Let δ_j denote the capping disk of c_j and let $\bar{X}_T = \bar{X} \cup ([0, T] \times ST^*S^3)$. Fix a sufficiently large $T > 0$ and replace $u(\partial\Sigma)$ in the construction of σ_u' above by the boundary of the chain $(u(\Sigma) \cap \bar{X}_T) \cup \bigcup_{j=1}^m \delta_j$ and then proceed as there. This means that we cap off the holomorphic curve by adding capping disks and construct a bounding chains of this capped disk.

5.3 Generalized holomorphic curves and the SFT-potential. The SFT counterpart of the chain map equation for augmentations is derived from 1-dimensional moduli spaces of generalized holomorphic curves. The moduli spaces are stratified and the key point of our construction is to patch the 1-dimensional strata in such a way that all boundary phenomena in the compact part of (X, L_K) cancel out, leaving only splitting at Reeb chords and intersections with bounding chains at infinity. We start by describing the curves in the 1-dimensional strata.

As in the disk case we assume we have a perturbation scheme for transversality. Again the perturbation is inductively constructed, we first perturb near the simplest curves (lowest energy and highest Euler characteristic) and then continue inductively in the hierarchy of curves, making all holomorphic curves transversely cut out and transverse with respect to the Morse data fixed. We also need transversality with respect to C_K that we explain next. A holomorphic curve u in general position has tangent vector along the boundary everywhere linearly independent of ∇f. Let the *shifting vector field* ν along ∂u be a vector field that together with the tangent vector of ∂u and ∇f gives a positively oriented triple. Let ∂u_ν denote ∂u shifted slightly along ν. By construction ∂u_ν is disjoint from a neighborhood of the boundary of σ_u. Let $u_{J\nu}$ denote u shifted slightly along an extension of $J\nu$ supported near the boundary of u. We chose the perturbation so that $u_{J\nu}$ is transverse to C_K.

With such perturbation scheme constructed we define generalized holomorphic curves to consist of the following data.

- A finite oriented graph Γ with vertex set V_Γ and edge set E_Γ.
- To each $v \in V_\Gamma$ is associated a (generic) holomorphic curve u^v with boundary on L_K (and possibly with positive punctures).
- To each edge $e \in E_\Gamma$ that has its endpoints at distinct vertices, $\partial e = v_+ - v_-$, $v_+ \neq v_-$, is associated an intersection point of the boundary curve ∂u^{v_-} and the bounding chain $\sigma_{u^{v_+}}$.
- To each edge $e \in E_\Gamma$ which has its endpoints at the same vertex v_0, $\partial e = v_0 - v_0 = 0$, is associated either an intersection point in $\partial u_\nu^{v_0} \cap \sigma_{u^{v_0}}$ or an intersection point in $u_{J\nu}^{v_0} \cap C_K$.

We call such a configuration a *generalized holomorphic curve over* Γ and denote it $\Gamma_{\mathbf{u}}$, where $\mathbf{u} = \{u_v\}_{v \in V_\Gamma}$ lists the curves at the vertices.

Remark 5.1. Several edges of a generalized holomorphic curve may have the same intersection point associated to them.

We define the Euler characteristic of a generalized holomorphic curve $\Gamma_{\mathbf{u}}$ as

$$\chi(\Gamma_{\mathbf{u}}) = \sum_{v \in V_\Gamma} \chi(u_v) - \#E_\Gamma,$$

where $\#E_\Gamma$ denotes the number of edges of Γ, and the dimension of the moduli space containing $\Gamma_{\mathbf{u}}$ as

$$\dim(\Gamma_{\mathbf{u}}) = \sum_{v \in V_\Gamma} \dim(u_v),$$

where $\dim(u_v)$ is the formal dimension of u_v.

In particular, if $\dim(\Gamma_{\mathbf{u}}) = 0$ then u_v is rigid for all $v \in V_\Gamma$ and if $\dim(\Gamma_{\mathbf{u}}) = 1$ then $\dim(u_v) = 1$ for exactly one $v \in V_\Gamma$ and u_v is rigid for all other $v \in V_\Gamma$. The relative homology class represented by $\Gamma_{\mathbf{u}}$ is the sum of the homology classes of the curves u_v at its vertices, $v \in V_\Gamma$.

We define the SFT-potential to be the generating function of generalized rigid curves over graphs Γ as just described:

$$\mathbf{F}_K = \sum_{m,k,\mathbf{c}^+} F_{m,k,\chi,\mathbf{c}^+}\, g_s^{-\chi+\ell(\mathbf{c}^+)}\, e^{mx} Q^k\, \mathbf{c}^+,$$

where $F_{m,k,\chi,\mathbf{c}^+}$ counts the algebraic number of generalized curves $\Gamma_{\mathbf{u}}$ in homology class $mx + kt \in H_2(X, L_K)$ with $\chi(\Gamma_{\mathbf{u}}) = \chi$ and with positive punctures according to the Reeb chord word $\mathbf{c}^+$. A generalized curve $\Gamma_{\mathbf{u}}$ contributes to this sum by the product of the weights of the curves at its vertices (the count coming from the perturbation scheme) times $\pm\frac{1}{2}$ for each edge where the sign is determined by the intersection number.

Remark 5.2. For computational purposes we note that we can rewrite the sum for $\mathbf{F}_K$ in a simpler way. Instead of the complicated oriented graphs with many edges considered above, we look at unoriented graphs with at most one edge connecting every pair of distinct vertices and no edge connecting a vertex to itself. We call such graphs *simple graphs*. We map complicated graphs to simple graphs by collapsing edges to the basic edge and removing self-edges. Then the contribution from all graphs lying over a simple graph is given the product of weights at the vertices times the product of $e^{\mathrm{lk}_e\, g_s}$, where the linking coefficient of an edge e connecting vertices corresponding to the curves u and u' is the intersection number $\sigma_u \cdot \partial u' = \partial u \cdot \sigma_{u'}$, and $e^{\frac{1}{2}\,\mathrm{slk}_v\, g_s}$, where the linking coefficient slk_v of a vertex v is the sum of intersection numbers $\partial u_\nu \cdot \sigma_u + u_{J\nu} \cdot C_K$, where u is the curve at v.

5.4 Compactification of 1-dimensional moduli spaces. The generalized holomorphic curves that we defined in Section 5.3 constitute the open strata of the 1-dimensional moduli. More precisely, the generalized curve $\Gamma_{\mathbf{u}}$ has a *generic* curve of dimension one at exactly one vertex. Except for the usual holomorphic degenerations in 1-parameter families, there are new boundary phenomena arising from the 1-dimensional curve becoming non-generic relative ∇f and C_K. More precisely we have the following description of the boundary of 1-dimensional starta of generalized holomorphic curves (we write u_v for the curve at vertex $v \in \Gamma_{\mathbf{u}}$).

Lemma 5.3. *Ekholm and Ng [n.d.] Generic degenerations of the holomorphic curves u_v at the vertices $v \in V_\Gamma$ are as follows (see Figure 3):*

(1) *Splitting at Reeb chords.*

(2) *Hyperbolic boundary splitting.*

(3) *Elliptic boundary splitting.*

Generic degenerations with respect to ∇f, C_K, and capping paths are as follows:

(4) *Crossing the stable manifold of κ_1: the boundary of the curve intersects the stable manifold of κ_1.*

(5) *Boundary crossing: a point in the boundary mapping to a bounding chain moves out across the boundary of a bounding chain.*

(6) *Interior crossing: An interior marked point mapping to C_K moves across the boundary L_K of C_K.*

(7) *Boundary kink: The boundary of a curve becomes tangent to ∇f at one point.*

(8) *Interior kink: A marked point mapping to C_K moves to the boundary in the holomorphic curve.*

(9) *The leading Fourier coefficient at a positive puncture vanishes.*

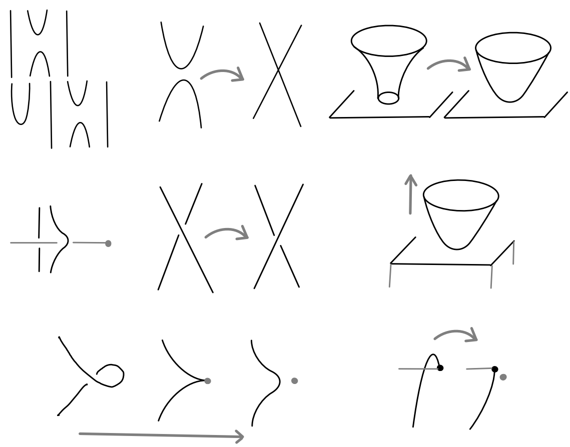


Figure 3: Degenerations in Lemma 5.3. Top row: (1), (2), (3), middle (4) (the dot is κ_1), (5), (6), bottom (7), (8) together, and (9) (gray dot represents $u_{J_\nu} \cap C_K$).

Proposition 5.4. *Boundaries of 1-dimensional strata of generalized holomorphic curves cancel out according to the following.*

(i) *The moduli space of generalized holomorphic curves does not change under degenerations* (4) *and* (9).

(ii) *Boundary splitting* (2) *cancel with boundary crossing* (5).

(iii) *Elliptic splitting* (3) *cancel with interior crossing* (6).

(iv) *Boundary kinks* (7) *cancel interior kinks* (8).

Proof. Consider (i). For (4), observe that as the boundary crosses the stable manifold of κ_1, the change in flow image is compensated by the change in the number of unstable manifold added. The invariance under (9) follows from a straightforward calculation using Fourier expansion near the Reeb chord: an intersection with the capping disk boundary turns into an intersection with C_K.

Consider (iv). A calculation in a local model for a generic tangency with ∇f shows that the self intersection of the boundary turns into an intersection with C_K. (This uses that the normal vector field of C_K is $\pm J\nabla f$.)

Consider (iii). Unlike (i) and (iv) this involves gluing holomorphic curves and therefore, as we will see, the details of the perturbation scheme (which also has further applications, see Ekholm and Shende [n.d.]).

At the hyperbolic boundary splitting we find a holomorphic curve with a double point that can be resolved in two ways, u_+ and u_-. Consider the two moduli spaces corresponding to m insertions at the corresponding intersection points between ∂u_+ and σ_{u_-} and ∂u_- and σ_{u_+}.

To obtain transversality at this singular curve for curves of any Euler characteristic we must separate the intersection points with the bounding chain. To this end, we use a perturbation scheme with multiple bounding chains that time-orders the boundary crossings. Each, now distinct, crossing can then be treated as a usual gluing. Consider gluing at m intersection points as ∂u_- crosses σ_{u_+}. This gives a curve of Euler characteristic decreased by m and orientation sign ϵ^m, $\epsilon = \pm 1$. Furthermore, at the gluing, the ordering permutation acts on the gluing strips and each intersection point is weighted by $\frac{1}{2}$. (The reason for the factor $\frac{1}{2}$ is that we count intersections between boundaries and bounding chains twice, for distinct curves both $\partial u \cap \sigma_v$ and $\partial v \cap \sigma_u$ contribute.) This gives a moduli space of additional weight

$$\epsilon^m \frac{1}{2^m m!} g_s^m.$$

The only difference between these configurations and those associated with the opposite crossing is the orientation sign. Hence the other gluing when ∂u_+ crosses σ_{u_-} gives the weight

$$(-1)^m \epsilon^m \frac{1}{2^m m!} g_s^m.$$

Noting that the original moduli space is oriented towards the crossing for one configuration and away from it for the other we find that the two gluings cancel if m is even and give a new curve of Euler characteristic decreased by m and of weight $\frac{2}{2^m m!}$ if m is odd. Counting ends of moduli spaces we find that the curves resulting from gluing at the crossing count with a factor

$$e^{\frac{1}{2}g_s} - e^{-\frac{1}{2}g_s}, \tag{5}$$

which cancels the change in linking number.

Cancellation (iii) follows from a gluing argument analogous to (ii): The curve with an interior point mapping to L_K can be resolved in two ways, one curve u^+ that intersects C_K at a point in the direction $+J\nabla f$ and one u^- that intersects C_K at a point in the direction $-J\nabla f$. A constant disk at the intersection point can be glued to the family of curves at the intersection with L_K. As in the hyperbolic case we separate the intersections and time order them to get transversality at any Euler characteristic. We then apply usual gluing and note that the intersection sign is part of the orientation data for the gluing problem, the calculation of weights is exactly as in the hyperbolic case above. (This time the $\frac{1}{2}$-factors comes from the boundary of C_K being twice L_K, $\partial C_K = 2[L_K]$.) We find again that glued configurations corresponds to multiplication by

$$e^{\frac{1}{2}g_s} - e^{-\frac{1}{2}g_s},$$

and cancels the difference in counts between $u^+_{J\nu} \cdot C_K$ and $u^-_{J\nu} \cdot C_K$. □

5.5 The SFT equation. We let $\mathbf{H}_K$ denote the count of generalized holomorphic curves $\Gamma_{\mathbf{u}}$, in $\mathbb{R} \times ST^*S^3$, rigid up to $\mathbb{R}$-translation. Such a generalized curve lies over a graph that has a main vertex corresponding to a curve of dimension 1, at all other vertices there are trivial Reeb chord strips. Consider such a generalized holomorphic curve $\Gamma_{\mathbf{u}}$. We write $\mathbf{c}^+(\mathbf{u})$ and $\mathbf{c}^-(\mathbf{u})$ for the monomials of positive and negative punctures of $\Gamma_{\mathbf{u}}$, write $w(\mathbf{u})$ for the weight of $\Gamma_{\mathbf{u}}$, $m(\mathbf{u})x + n(\mathbf{u})p + l(\mathbf{u})t$ for its homology class, and $\chi(\mathbf{u})$ for the Euler characteristic of the generalized curve of $\Gamma_{\mathbf{u}}$. Define the SFT-Hamiltonian

$$\mathbf{H}_K = \sum_{\dim(\Gamma_{\mathbf{u}})=1} w(\mathbf{u})\, g_s^{-\chi(\mathbf{u})+\ell(\mathbf{c}^+(\mathbf{u}))}\, e^{m(\mathbf{u})x+n(\mathbf{u})p+l(\mathbf{u})t}\, \mathbf{c}^+(\mathbf{u})\, \partial_{\mathbf{c}^-(\mathbf{u})},$$

where the sum ranges over all generalized holomorphic curves. As above this formula can be simplified to a sum over simpler graphs with more elaborate weights on edges.

Lemma 5.5. *Consider a curve u at infinity in class $mx + np + kt$. The count of the corresponding generalized curves with insertion along ∂u equals*

$$e^{-\mathbf{F}_K} e^{mx} Q^k e^{n g_s \frac{\partial}{\partial x}} e^{\mathbf{F}_K}.$$

Proof. Contributions from bounding chains of curves inserted r times along np corresponds to multiplication by

$$n^r \frac{1}{r!} g_s^{-r} \sum_{r_1+\cdots+r_j=r} \frac{\partial^{r_1}\mathbf{F}_K}{\partial x^{r_1}} \cdots \frac{\partial^{r_j}\mathbf{F}_K}{\partial x^{r_j}},$$

where a factor $\frac{\partial^s \mathbf{F}_K}{\partial x^s}$ corresponds to attaching the bounding chain of a curve s times. □

Theorem 5.6. *If K is a knot and $L_K \subset X$ its conormal Lagrangian then the SFT equation*

$$e^{-\mathbf{F}_K}\, \mathbf{H}_K|_{p=g_s \frac{\partial}{\partial x}}\, e^{\mathbf{F}_K} = 0 \tag{6}$$

holds.

Proof. Lemma 5.3, Proposition 5.4, and Lemma 5.5 show that the terms in left hand side of (6) counts the ends of a compact oriented 1-dimensional moduli space. □

Remark 5.7. We point out that Lemma 5.5 gives an enumerative geometrical meaning to the standard quantization scheme $p = g_s \frac{\partial}{\partial x}$ by counting insertions of bounding chains. See Ekholm [2014, Section 3.3] for a related path integral argument.

5.6 Framing and Gromov–Witten invariants.

Lemma 5.3 and Proposition 5.4 imply that the open Gromov–Witten potential of L_K is invariant under deformation. Recall from Section 1.2 that dualities between string and gauge theories imply that

$$\Psi_K(x,Q) = e^{F_K(x,Q)} = \sum_m H_m(e^{g_s},Q)e^{mx},$$

where H_m is the m-colored HOMFLY-PT polynomial. It is well-known that the colored HOMFLY-PT polynomial depends on framing. We derive this dependence here using our definition of generalized holomorphic curves. Assume that Ψ_K above is defined for a framing (x,p) of Λ_K. Then other framings are given by $(x',p') = (x+rp,p)$ where r is an integer. Let $\Psi^r_K(x',Q)$ denote the wave function defined using the framing (x',p').

Theorem 5.8. *If $\Psi_K(x,Q)$ is as above then*

$$\Psi^r_K(x',Q) = \sum_m H_m(e^{g_s},Q)\, e^{m^2 r g_s} e^{mx'}.$$

Proof. Note first that the actual holomorphic curves are independent of the framing. The change thus comes from the bounding chains: the boundaries at infinity $\partial_\infty \sigma_u$ must be corrected to lie in multiples of the new preferred class x'. Thus, for a curve that goes m times around the generator of $H_1(L_K)$, we must correct the bounding chain adapted to x by adding $mrW^{\mathrm{u}}(\kappa_1)$. Under such a change, the linking number in L_K in this class changes by $m^2 r$. □

5.7 Quantization of the augmentation variety in basic examples.

5.7.1 The unknot. Using Morse flow trees it is easy to see that there are no higher genus curves with boundary on Λ_U. As with the augmentation polynomial, there are no additional operators to eliminate for the unknot and $\mathbf{H}_U$ gives the operator equation directly:

$$\widehat{\mathrm{Aug}}_U = 1 - e^{\hat{x}} - e^{\hat{p}} - Qe^{\hat{x}}e^{\hat{p}},$$

which agrees with the the recursion relation for the colored HOMFLY-PT, see e.g. Aganagic and Vafa [2012].

5.7.2 The trefoil. It can be shown Ekholm and Ng [n.d.] that there are no higher genus curves with boundary on Λ_T. The SFT Hamiltonian can again be computed from disks with flow lines attached. If c is a chord with $|c| = 1$, we write $H(c)$ for the part of the Hamiltonian $\mathbf{H}_T$ with a positive puncture at c and leave out c from the notation. Then relevant parts of the Hamiltonian are:

$$\begin{aligned}
H(b_{12}) &= e^{-\hat{x}}\partial_{a_{12}} - \partial_{a_{21}} + \mathcal{O}(a)\\
H(c_{11}) &= e^{\hat{x}}e^{\hat{p}} - e^{-g_s}e^{\hat{x}} - ((1+e^{-g_s})Q - e^{\hat{p}})\partial_{a_{12}} - Q\partial^2_{a_{12}}\partial_{a_{21}} + \mathcal{O}(a)\\
H(c_{21}) &= Q - e^{\hat{p}} + e^{\hat{x}}e^{\hat{p}}\partial_{a_{21}} + Q\partial_{a_{12}}\partial_{a_{21}} + (e^{-g_s}-1)e^{\hat{x}}a_{12}\\
&\qquad + (e^{-g_s}-1)Qa_{12}\partial_{a_{12}} + \mathcal{O}(a^2)\\
H(c_{22}) &= e^{\hat{p}} - 1 - Q\partial_{a_{21}} + e^{\hat{p}}\partial_{a_{12}}\partial_{a_{21}} + (e^{g_s}-1)Qa_{12}\\
&\qquad + (e^{g_s}-1)e^{\hat{p}}a_{12}\partial_{a_{12}} + \mathcal{O}(a^2),
\end{aligned}$$

where $\mathcal{O}(a)$ represents order in the variables $a = (a_{12}, a_{21})$. The factors $(e^{g_s} - 1)$ in front of disks with additional positive punctures comes from the perturbation scheme and are related to the gluing analysis in the proof of Proposition 5.4, see Ekholm and Ng [ibid.]. In close analogy with the calculation at the classical level, the operators $\partial_{a_{12}}$ and $\partial_{a_{21}}$ can be eliminated and we get an operator equation which after change of framing to make x correspond to the longitude of T, i.e., 0-framing, becomes

$$\begin{aligned}
\widehat{\mathrm{Aug}}_T &= e^{g_s}Q^3e^{3\hat{p}}(Q - e^{-3g_s}e^{2\hat{p}})(Q - e^{-g_s}e^{\hat{p}})\cdot 1\\
&\quad + e^{-5g_s/2}(Q - e^{-2g_s}e^{2\hat{p}})\Big((e^{2g_s}e^{2\hat{p}} + e^{3g_s}e^{2\hat{p}} - e^{3g_s}e^{\hat{p}} + e^{4g_s})Q^2\\
&\qquad\qquad - (e^{g_s}e^{3\hat{p}} + e^{3g_s}e^{2\hat{p}} + e^{g_s}e^{2\hat{p}})Q + e^{4\hat{p}}\Big)\cdot e^{\hat{x}}\\
&\quad + (Q - e^{-g_s}e^{2\hat{p}})(e^{\hat{p}} - e^{g_s})\cdot e^{2\hat{x}},
\end{aligned}$$

in agreement with the recursion relation of the colored HOMFLY-PT in Garoufalidis, Lauda, and Le [2016].

References

Mina Aganagic, Tobias Ekholm, Lenhard Ng, and Cumrun Vafa (2014). "Topological strings, D-model, and knot contact homology". *Adv. Theor. Math. Phys.* 18.4, pp. 827–956. MR: 3277674 (cit. on p. 1084).

Mina Aganagic and Cumrun Vafa (2012). "Large N Duality, Mirror Symmetry, and a Q-deformed A-polynomial for Knots". arXiv: 1204.4709 (cit. on pp. 1083, 1104).

Kai Cieliebak, Tobias Ekholm, Janko Latschev, and Lenhard Ng (2017). "Knot contact homology, string topology, and the cord algebra". *J. Éc. polytech. Math.* 4, pp. 661–780. MR: 3665612 (cit. on pp. 1090, 1091).

T. Ekholm (2002). "The complex shade of a real space, and its applications". *Algebra i Analiz* 14.2, pp. 56–91. MR: 1925881 (cit. on p. 1096).

Tobias Ekholm (2007). "Morse flow trees and Legendrian contact homology in 1-jet spaces". *Geom. Topol.* 11, pp. 1083–1224. arXiv: math.SG/0509386 (cit. on p. 1088).

– (2014). "Notes on topological strings and knot contact homology". In: *Proceedings of the Gökova Geometry-Topology Conference 2013*. Gökova Geometry/Topology Conference (GGT), Gökova, pp. 1–32. MR: 3287796 (cit. on p. 1103).

Tobias Ekholm, John B. Etnyre, Lenhard Ng, and Michael G. Sullivan (2013). "Knot contact homology". *Geom. Topol.* 17.2, pp. 975–1112. MR: 3070519 (cit. on pp. 1082, 1085, 1088).

Tobias Ekholm, John Etnyre, Lenhard Ng, and Michael Sullivan (2013). "Filtrations on the knot contact homology of transverse knots". *Math. Ann.* 355.4, pp. 1561–1591. MR: 3037024 (cit. on pp. 1082, 1088).

Tobias Ekholm, John Etnyre, and Michael Sullivan (2007). "Legendrian contact homology in $P \times \mathbb{R}$". *Trans. Amer. Math. Soc.* 359.7, pp. 3301–3335. arXiv: math/0505451. MR: 2299457 (cit. on p. 1087).

Tobias Ekholm, Ko Honda, and Tamás Kálmán (2016). "Legendrian knots and exact Lagrangian cobordisms". *J. Eur. Math. Soc. (JEMS)* 18.11, pp. 2627–2689. MR: 3562353 (cit. on p. 1093).

Tobias Ekholm and Yanki Lekili (2017). "Duality between Lagrangian and Legendrian invariants". arXiv: 1701.01284 (cit. on pp. 1087, 1092).

Tobias Ekholm and Lenhard Ng (n.d.). "Higher genus knot contact homology and recursion for colored HOMFLY-PT polynomials". arXiv: 1803:04011 (cit. on pp. 1096, 1097, 1099, 1104).

Tobias Ekholm, Lenhard Ng, and Vivek Shende (Oct. 2017a). "A complete knot invariant from contact homology". *Inventiones mathematicae* (cit. on pp. 1082, 1089–1091).

– (2017b). "A complete knot invariant from contact homology". arXiv: 1606.07050 (cit. on p. 1092).

Tobias Ekholm and Vivek Shende (n.d.). “Holomorphic curves and the skein relation”. In preparation (cit. on p. 1101).

Y. Eliashberg, A. Givental, and H. Hofer (2000). “Introduction to symplectic field theory”. *Geom. Funct. Anal.* Special Volume, Part II. GAFA 2000 (Tel Aviv, 1999), pp. 560–673. arXiv: math.SG/0010059 (cit. on p. 1082).

Yakov Eliashberg (2007). “Symplectic field theory and its applications”. In: *International Congress of Mathematicians. Vol. I.* Eur. Math. Soc., Zürich, pp. 217–246. MR: 2334192 (cit. on p. 1082).

Kenji Fukaya, Yong-Geun Oh, Hiroshi Ohta, and Kaoru Ono (2009). *Lagrangian intersection Floer theory: anomaly and obstruction. Part I*. Vol. 46. AMS/IP Studies in Advanced Mathematics. Providence, RI: American Mathematical Society, pp. xii+396 (cit. on p. 1094).

Sheel Ganatra, John Pardon, and Vivek Shende (2017). “Covariantly functorial Floer theory on Liouville sectors”. arXiv: 1706.03152 (cit. on p. 1092).

Stavros Garoufalidis, Aaron D. Lauda, and Thang T.Q. Le (2016). “The colored HOMFLY polynomial is q-holonomic”. arXiv: 1604.08502 (cit. on pp. 1083, 1104).

Vito Iacovino (2009a). “Open Gromov-Witten theory on Calabi-Yau three-folds I”. arXiv: 0907.5225 (cit. on p. 1093).

– (2009b). “Open Gromov-Witten theory on Calabi-Yau three-folds II”. arXiv: 0908.0393 (cit. on p. 1093).

Sergiy Koshkin (2007). “Conormal bundles to knots and the Gopakumar-Vafa conjecture”. *Adv. Theor. Math. Phys.* 11.4, pp. 591–634. MR: 2354076 (cit. on pp. 1083, 1092).

Lenhard Ng (2008). “Framed knot contact homology”. *Duke Math. J.* 141.2, pp. 365–406. arXiv: math/0407071. MR: 2376818 (cit. on p. 1082).

– (2011). “Combinatorial knot contact homology and transverse knots”. *Adv. Math.* 227.6, pp. 2189–2219. arXiv: 1010.0451. MR: 2807087 (cit. on p. 1082).

Hirosi Ooguri and Cumrun Vafa (2000). “Knot invariants and topological strings”. *Nucl. Phys.* B577, pp. 419–438. arXiv: hep-th/9912123 (cit. on p. 1082).

– (2002). “World sheet derivation of a large N duality”. *Nucl. Phys.* B641, pp. 3–34. arXiv: hep-th/0205297 (cit. on p. 1082).

Vivek Shende (2016). “The conormal torus is a complete knot invariant”. arXiv: 1604.03520 (cit. on pp. 1082, 1092).

Zachary Sylvan (2016). “On partially wrapped Fukaya categories”. arXiv: 1604.02540 (cit. on p. 1092).

Oleg Viro (2001). “Encomplexing the writhe”. In: *Topology, ergodic theory, real algebraic geometry*. Vol. 202. Amer. Math. Soc. Transl. Ser. 2. Amer. Math. Soc., Providence, RI, pp. 241–256. MR: 1819192 (cit. on p. 1096).

Edward Witten (1995). “Chern-Simons gauge theory as a string theory”. *Prog.Math.* 133, pp. 637–678. arXiv: hep-th/9207094 (cit. on p. 1082).

Received 2017-11-16.

Tobias Ekholm
Department of mathematics, Uppsala University, Box 480, 751 06 Uppsala, Sweden
and
Institut Mittag-Leffler, Aurav 17, 182 60 Djursholm, Sweden
tobias.ekholm@math.uu.se
ekholm@mittag-leffler.se

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1105–1132)

CONSTRUCTING GROUP ACTIONS ON QUASI-TREES

Koji Fujiwara (藤原　耕二)

Abstract

A quasi-tree is a geodesic metric space quasi-isometric to a tree. We give a general construction of many actions of groups on quasi-trees. The groups we can handle include non-elementary hyperbolic groups, CAT(0) groups with rank 1 elements, mapping class groups and the outer automorphism groups of free groups. As an application, we show that mapping class groups act on finite products of Gromov-hyperbolic spaces so that orbit maps are quasi-isometric embeddings. It implies that mapping class groups have finite asymptotic dimension.

1 Introduction

1.1 Overview. Group actions are useful in the study of infinite (discrete) groups. One example is the theory of groups acting on simplicial trees by automorphisms, called *Bass-Serre theory*, Serre [1980]. Serre observed that $SL(2,\mathbb{Z})$ properly and co-compactly acts on an infinite simplicial tree, which is embedded in the upper half plane. On the other hand he proved that if $SL(3,\mathbb{Z})$ acts on any simplicial tree by automorphisms then there is a fixed point. Using the theory, he obtained a geometric proof of the theorem by Ihara saying that every torsion-free discrete subgroup of $SL(2,\mathbb{Q}_p)$ is free.

A central idea in Geometric group theory is to use *hyperbolicity* (in the sense of Gromov) of a space to prove algebraic properties of a group that acts on it. A tree is a most elementary example of a hyperbolic space. This method created the theory of *hyperbolic groups*, Gromov [1987]. Another example is an approach by Masur-Minsky to the mapping class group of a surface using the hyperbolicity of the curve complex of the surface. We will rely on their theory for our application.

This note is a survey of the work by Bestvina, Bromberg, and Fujiwara [2015], which uses *quasi-trees* to study groups. A quasi-tree is a geodesic metric space that is *quasi-isometric* to (ie, "looks like", see the precise definition later) a simplicial tree. A quasi-tree

I would like to thank my parents. This work is supported by KAKENHI (23244005, 15H05739).
MSC2010: 20F65.
Keywords: quasi-trees, projection complex, mapping class groups, hyperbolic groups, asymptotic dimension.

is always hyperbolic in the sense of Gromov. There are many advantages in this approach. One is that quasi-trees are more flexible than trees, so that there are in fact more groups that act on quasi-trees than on trees. Since quasi-trees are hyperbolic, many techniques and results that are obtained for groups acting on hyperbolic spaces apply, but moreover, we sometimes obtain stronger conclusions since quasi-trees are special and easier to handle.

Also, we introduce new methods to produce quasi-trees, called *projection complex*, equipped with an isometric group action by a given group. For that we only need to check a small set of axioms, which are satisfied by many examples (see Examples 4.3), including *hyperbolic groups*, mapping class groups and the outer automorphism group $Out(F_n)$ of a free group F_n of rank n, and are able to produce many actions that are sometimes hidden at a first glance. We also construct closely related space called *quasi-tree of metric spaces*. Using those constructions we prove new theorems and also recover some known ones.

1.2 Intuitive description of the main construction. To explain the idea by an example, consider a discrete group Γ of isometries of hyperbolic n-space $\mathbb{H}^n$ and let $\gamma \in \Gamma$ be an element with an axis (ie, a γ-invariant geodesic that γ acts on by a translation) $\ell \subset \mathbb{H}^n$. Denote by $\mathbf{Y}$ the set of all Γ-translates of ℓ, i.e. the set of axes of conjugates of γ. Now we will construct a quasi-tree Q with a G-action from the disjoint union of the translates of ℓ by joining pairs of translates by edges following a certain rule. We want: the resulting space Q is connected; Q looks like a tree, so Q should not contain larger and larger embedded circles (then Q is not a quasi-tree); so that we put edges as few as possible just enough to make Q to be connected; the connecting rule is G-equivariant to have a G action on Q. Intuitively, the rule we will use is simple. For elements (ie, lines) $A, B \in \mathbf{Y}$ we do not want to join them when there is another element C that is "between A and B" (then we would rather join A and C, and also C and B). So, we join A and B by an edge only when there is no element C between A and B.

But how do we define that "C is between A and B"? When $A, C \in \mathbf{Y}$, $A \neq C$, denote by $\pi_C(A) \subset C$ the image of A under the nearest point projection $\pi_C : \mathbb{H}^n \to C$. We call this set the *projection* of A to C and we observe:

(P0) The diameter $\operatorname{diam} \pi_C(A)$ is uniformly bounded by some constant $\theta \geq 0$, independently of $A, C \in \mathbf{Y}$.

This is a consequence of discreteness of G, because a line in $\mathbb{H}^n$ will have a big projection to another line only if the two lines have long segments with small Hausdorff distance between them, so that there is a uniform upper bound unless they coincide, since G is discrete. When $B \neq A \neq C$ we define a "distance" by

$$d_C^{\pi}(A, B) = \operatorname{diam}(\pi_C(A) \cup \pi_C(B)).$$

Now fix a constant $K >> \theta$. We say C is between A and B if $d^{\pi}_{C}(A, B) \geq K$. (In fact we slightly perturb the distance functions d^{π}_{C} at once in advance.) Now the rule is that if there is no such $C \in \mathbf{Y}$ then we join A and B by an edge connecting $\pi_A(B)$ and $\pi_B(A)$, which are small (imagine θ is small) sets by (P0). The resulting space, $\mathcal{C}(\mathbf{Y})$, turns out to be a quasi-tree.

One may think this example is special since each A is a line. Of course, $\mathcal{C}(\mathbf{Y})$ would not be a quasi-tree if $A \in \mathbf{Y}$ were not a quasi-tree since A is embedded in $\mathcal{C}(\mathbf{Y})$. By collapsing each A to a point in $\mathcal{C}(\mathbf{Y})$ we obtain a new space $\mathcal{P}(\mathbf{Y})$ (then the geometry of A becomes irrelevant), which is again a quasi-tree. Our discovery is that if we start with a given abstract data: a set $\mathbf{Y}$ and maps(projections) between any two elements in $\mathbf{Y}$ satisfying a small set of "Axioms", and construct a space $\mathcal{P}(\mathbf{Y})$ by the rule we explained, $\mathcal{P}(\mathbf{Y})$ is always a quasi-tree. It does not matter how we obtain the data as long as it satisfies the axioms.

The technical difficulty in the construction is to perturb the distance functions d^{π}_{A} by a bounded amount as we said (see Section 2.1 for details). Without this perturbation the resulting space may contain larger and larger loops and is not a quasi-tree even if the initial data satisfies the axioms (see Bestvina, Bromberg, and Fujiwara [2015] for a counter example).

1.3 Axioms for quasi-tree of metric spaces and projection complex. We continue with the example, and explain the Axioms behind the construction. First, the distance function d^{π}_{Y} is always symmetric and satisfies the triangle inequality (nothing to do with discreteness of G):

(P3) $d^{\pi}_{Y}(X, Z) = d^{\pi}_{Y}(Z, X)$;

(P4) $d^{\pi}_{Y}(X, Z) + d^{\pi}_{Y}(Z, W) \geq d^{\pi}_{Y}(X, W)$,

but in general we have $d^{\pi}_{A}(B, B) > 0$, so this is a pseudo-distance function (we frequently drop "pseudo") . We observe further, again since Γ is discrete, for a perhaps larger constant θ:

(P1) For any triple $A, B, C \in \mathbf{Y}$ of distinct elements, at most one of the following three numbers is greater than θ:

$$d^{\pi}_{A}(B, C), d^{\pi}_{B}(A, C), d^{\pi}_{C}(A, B).$$

(P2) For any $A, B \in \mathbf{Y}$ the following set is finite:

$$\{C \in \mathbf{Y} \mid d^{\pi}_{C}(A, B) > \theta\}.$$

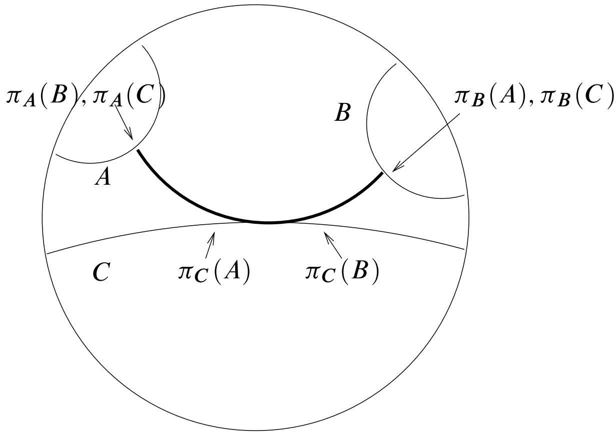


Figure 1: Axiom (P1). The bold line is the shortest segment between A and B in $\mathbb{H}^n$. Note that C and this segment stay close for a long time, therefore $d_C^\pi(A, B)$ is large, while $d_A^\pi(B, C)$ and $d_B^\pi(A, C)$ are small. (Figure from Bestvina, Bromberg, and Fujiwara [2015])

For an even more basic example where (P0)-(P2) hold with $\theta = 0$, consider the Cayley tree (Cayley graph) of the free group $F_2 = \langle a, b \rangle$ and for $\mathbf{Y}$ take the F_2-orbit of the axis of a.

We can show that (P0), (P1) and (P2) are sufficient *Axioms* for the space we construct to be a quasi-tree. We start with a collection of metric spaces $\mathbf{Y}$ and a collection of subsets $\pi_A(B) \subset A$ for $A \neq B$ with d_A^π satisfying (P0)-(P2). Note that (P3) and (P4) always hold for d_A^π. Then we construct a space by putting edges between (disjoint union of) the metric spaces in $\mathbf{Y}$ following our rule. Here is a summary theorem. (1) applies to the $\mathbb{H}^n$ example. In a way we "reconstruct" the ambient space with the group action from $\mathbf{Y}$. But the punch line is that we do not have to start with an ambient space and its subspaces in the theorem.

Theorem 1.1. *Bestvina, Bromberg, and Fujiwara [2015] Suppose $\mathbf{Y}$ is a collection of geodesic metric spaces and for every $A, B \in \mathbf{Y}$ with $A \neq B$ we are given a subset $\pi_A(B) \subset A$ such that (P0)-(P2) hold for the distance functions d_A^π for a constant θ.*

Then there is a geodesic metric space $\mathcal{C}(\mathbf{Y})$ that contains isometrically embedded, totally geodesic, pairwise disjoint copies of each $A \in \mathbf{Y}$ such that for all $A \neq B$ the nearest point projection of B to A in $\mathcal{C}(\mathbf{Y})$ is a uniformly bounded set uniformly close to $\pi_A(B)$ such that

(1) *The construction is equivariant, namely, if a group G acts isometrically on the disjoint union of the spaces in $\mathbf{Y}$ preserving projections, i.e., $g(\pi_A(B)) = \pi_{gA}(gB)$ for any $A, B \in \mathbf{Y}$ and $g \in G$, then the group action extends to $\mathcal{C}(\mathbf{Y})$.*

(2) *The quotient $\mathcal{C}(\mathbf{Y})/\mathbf{Y}$ obtained by collapsing the embedded copies of each $Y \in \mathbf{Y}$ to a point is a quasi-tree.*

Some explanations are in order. A subset $\mathcal{Y} \subset \mathcal{X}$ is *totally geodesic* if any geodesic in $\mathcal{X}$ joining two points in $\mathcal{Y}$ is contained in $\mathcal{Y}$. The space $\mathcal{C}(\mathbf{Y})$ is called a *quasi-tree of metric spaces*. Its construction will depend on the choice of a sufficiently large parameter $K >> \theta$, and it would be more precise to denote the space by $\mathcal{C}_K(\mathbf{Y})$. If $K < K'$ there is a natural Lipschitz map $\mathcal{C}_K(\mathbf{Y}) \to \mathcal{C}_{K'}(\mathbf{Y})$ which is in general not a quasi-isometry, and in fact unbounded sets may map to bounded sets.

The space $\mathcal{C}(\mathbf{Y})/\mathbf{Y}$ is called the *projection complex*, denoted by $\mathcal{P}(\mathbf{Y}) = \mathcal{P}_K(\mathbf{Y})$, which depends on K. We can think of the quasi-tree of metric spaces $\mathcal{C}(\mathbf{Y})$ as being obtained from $\mathcal{P}(\mathbf{Y})$, which is a quasi-tree by (2), by blowing up vertices to corresponding metric spaces. This explains the terminology.

Many properties that hold uniformly for the spaces in $\mathbf{Y}$ carry over to $\mathcal{C}(\mathbf{Y})$. For example (there will be the item (iii) later):

Theorem 1.2. *Bestvina, Bromberg, and Fujiwara [ibid.]. Let $\mathcal{C}(\mathbf{Y})$ be the quasi-tree of metric spaces $\mathbf{Y}$ constructed in Theorem 1.1.*

(i) *If each $X \in \mathbf{Y}$ is isometric to $\mathbb{R}$ then $\mathcal{C}(\mathbf{Y})$ is a quasi-tree; more generally, if all $X \in \mathbf{Y}$ are quasi-trees with a uniform bottleneck constant then $\mathcal{C}(\mathbf{Y})$ is a quasi-tree.*

(ii) *If each $X \in \mathbf{Y}$ is δ-hyperbolic with the same δ, then $\mathcal{C}(\mathbf{Y})$ is hyperbolic.*

Here, a geodesic metric space X satisfies the *bottleneck property* if there exists $\Delta \geq 0$ such that for any two points $x, y \in X$ the midpoint z of a geodesic between x and y satisfies the property such that any path from x to y intersects the Δ-ball centered at z. The constant Δ is called the *bottleneck constant*. Manning [2005] showed that X satisfying the bottleneck property is equivalent to X being a quasi-tree. Note that (i) in particular says that the space $\mathcal{C}(\mathbf{Y})$ obtained from an orbit of axes in $\mathbb{H}^n$ as in the example is a quasi-tree and not (quasi-isometric to) $\mathbb{H}^n$.

1.4 Asymptotic dimension.

We give another example of a property that descends from spaces in $\mathbf{Y}$ to $\mathcal{C}(\mathbf{Y})$. The notion of *asymptotic dimension* was introduced by Gromov Gromov [1993] as a large-scale analog of the covering dimension.

Definition 1.3 (Asymptotic dimension). A metric space X has *asymptotic dimension* $\operatorname{asdim}(X) \leq n$ if for every $R > 0$ there is a covering of X by uniformly bounded sets such that every metric R-ball intersects at most $n+1$ of the sets in the cover. More generally, a collection of metric spaces has asdim at most n *uniformly* if for every R there are covers of each space as above whose elements are uniformly bounded over the whole collection.

In **Theorem 1.2** we also have:

(iii) *If the collection* $\mathbf{Y}$ *has* asdim $\leq n$ *uniformly, then* $\text{asdim}(\mathcal{C}(\mathbf{Y})) \leq n+1$.

As an application we have:

Theorem 1.4. *Bestvina, Bromberg, and Fujiwara [2015]. Let Σ be a closed orientable surface, possibly with punctures, and $MCG(\Sigma)$ its mapping class group. Then* $\text{asdim}(MCG(\Sigma)) < \infty$.

The *Coarse Baum-Connes conjecture* (for torsion free subgroups of finite index) and therefore the *Novikov conjecture* for $MCG(\Sigma)$ follows from Theorem 1.4 by a work of Yu [1998], cf. Roe [2003]. Various other statements that imply the Novikov conjecture for $MCG(\Sigma)$ were known earlier (see Kida [2008], Hamenstädt [2009], and J. A. Behrstock and Minsky [2008]).

1.5 Some basic notions. We collect some standard definitions we use.

Let X, Y be two metric spaces. We often denote the distance between x, y by $|x-y|$. A map $f : X \to Y$ is a (K, L)-*quasi-isometric embedding* if for all points $x, y \in X$,

$$\frac{|x-y|}{K} - L \leq |f(x) - f(y)| \leq K|x-y| + L.$$

If it additionally satisfies that for all point $y \in Y$ there exists $x \in X$ such that $|y - f(x)| \leq L$, then we say f is a *quasi-isometry*, and X and Y are *quasi-isometric*. In those definitions, only the existence of constants K, L is important, and we sometimes omit them. Quasi-isometry is an equivalence relation among metric spaces.

Suppose a group G acts on a metric space X by isometries. We say the action is *co-compact/co-bounded* if the quotient is compact/bounded. We say the action is *proper* (or, *properly discontinuous*) if for any $R > 0$ and $x \in X$ the number of elements $g \in G$ with $|x - gx| \leq R$ is finitely many.

Let G be a finitely generated group and S a finite set of generators. We assume that if $s \in S$ then $s^{-1} \in S$. We form a graph as follows: there is a vertex for each element $g \in G$. We join two vertices $g, h \in G$ if there is $s \in S$ with $h = gs$. The graph is called a *Cayley graph* of G, denoted by $Cay(G, S)$. Since S generates G, the graph $Cay(G, S)$ is connected. G acts on the graph by automorphisms from the left: An element $g \in G$ sends a vertex $h \in G$ to a vertex $gh \in G$. By declaring each edge has length 1, $Cay(G, S)$ becomes a geodesic space. The action by G is proper and co-compact. The distance between the identity and $g \in G$ is denoted by $|g|$ and called the *word norm* of g.

Let $\Delta(a,b,c)$ be a geodesic triangle in the hyperbolic plane $\mathbb{H}^2$, where the three sides a,b,c are geodesics. Gauss-Bonnet theorem says that the area of Δ is at most π. It then follows that each side is contained in the 2-neighborhood of the union of the other two sides. Gromov turned this uniform thinness of geodesic triangles into a definition. A geodesic metric space X is *δ-hyperbolic* for a constant δ if all geodesic triangle in X is *δ-thin*, namely, each side is contained in the δ-neighborhood of the union of the other two. We often suppress δ and say X is *hyperbolic*. The hyperbolic spaces $\mathbb{H}^n$ are hyperbolic, trees are hyperbolic, but the Euclidean plane is not hyperbolic. A finitely generated group is a *(word) hyperbolic group* if it acts on a hyperbolic space properly and co-compactly by isometries, equivalently, if its Cayley graph is hyperbolic. If G contains $\mathbb{Z}^2$ then it is not hyperbolic. Another important class of spaces is of *CAT(0) spaces* (or *Hadamard spaces*), which are, roughly speaking, complete, simply connected, and "non-positively curved" geodesic spaces. See for example Ballmann [1995]. This class is a good source of examples.

The *translation length* $\tau(g)$ of an isometry $g : X \to X$ of a metric space X is

$$\tau(g) := \lim_{k\to\infty} \frac{d_X(x, g^k(x))}{k}.$$

The limit exists and is independent of $x \in X$. We say the isometry is *hyperbolic* if $\tau(g) > 0$.

We organize the rest of this note as follows: in Section 2 we define projection complex and quasi-tree of metric spaces from the beginning, which is independent from Section 1 (so that there is an overlap). In Section 3 we discuss application to mapping class groups. In Section 4 we give many examples that satisfy the axioms, and also discuss other applications.

Acknowledgments. I would like to thank Mladen Bestvina and Kenneth Bromberg. Most of the work presented here is from a long collaboration with them. I am grateful to Bestvina for reading a draft and giving useful suggestions.

2 Definition of projection complex and quasi-tree of metric spaces

We start over and will give a precise setting and conditions for our construction, Bestvina, Bromberg, and Fujiwara [2015]. To define the projection complex, we do not really need the projections $\pi_A(B)$ as in the example on $\mathbb{H}^n$; we only need the pseudo-distances $d_C^\pi(A,B)$.

2.1 Projection complex axioms. Let $\mathbf{Y}$ be a set, and assume that for each $Y \in \mathbf{Y}$ we have a function

$$d_Y^\pi : (\mathbf{Y} \setminus \{Y\}) \times (\mathbf{Y} \setminus \{Y\}) \longrightarrow [0, \infty).$$

Let $\theta \geq 0$ be a constant. Assume the following (PC1) - (PC4) are satisfied (they are same as (P1)-(P4) except for the order). We call them *projection complex axioms*.

(PC 1) $d_Y^\pi(X, Z) = d_Y^\pi(Z, X)$ for all distinct X, Y, Z;

(PC 2) $d_Y^\pi(X, Z) + d_Y^\pi(Z, W) \geq d_Y^\pi(X, W)$ for all distinct X, Y, Z, W (triangle inequality);

(PC 3) $\min\{d_Y^\pi(X, Z), d_Z^\pi(X, Y)\} \leq \theta$ for all distinct X, Y, Z;

(PC 4) for all $X, Z \in \mathbf{Y}$, $\#\{Y | d_Y^\pi(X, Z) > \theta\}$ is finite.

As an analog of (P0), uniform boundedness of the projections $\pi_Y(Z)$, we could require (PC0), but this will not be used to define a projection complex.

(PC 0) $d_Y^\pi(Z, Z) \leq \theta$ for all distinct Y, Z.

Before we define the projection complex, there is one technical difficulty we have to deal with. Given distance functions that satisfy (PC1) - (PC4), we modify them by a bounded amount for our purpose. This modification is a key to define an order on a set $\mathbf{Y}_K(X, Z)$ we define later.

For $X, Z \in \mathbf{Y}$ with $X \neq Z$ let $\mathcal{H}(X, Z)$ be the set of pairs $(X', Z') \in \mathbf{Y} \times \mathbf{Y}$ with $X' \neq Z'$ such that one of the following four holds:

- both $d_X^\pi(X', Z'), d_Z^\pi(X', Z') > 2\theta$;
- $X = X'$ and $d_Z^\pi(X, Z') > 2\theta$;
- $Z = Z'$ and $d_X^\pi(X', Z) > 2\theta$;
- $(X', Z') = (X, Z)$.

We then define the modified distance functions

$$d_Y : (\mathbf{Y} \setminus \{Y\}) \times (\mathbf{Y} \setminus \{Y\}) \to [0, \infty)$$

by $d_Y(X, Z) = 0$ if Y is contained in a pair in $\mathcal{H}(X, Z)$, and otherwise,

$$d_Y(X, Z) = \inf_{(X', Z') \in \mathcal{H}(X, Z)} d_Y^\pi(X', Z').$$

For example, if $d^\pi_Y(W,Z) > 2\theta$, then $(W,Z) \in \mathcal{H}(Y,Z)$ and $d_W(Y,Z) = 0$. Note that it is clear from the definition that $d_Y(X,Z) \le d^\pi_Y(X,Z)$ and therefore (PC3) and (PC4) still hold for d_Y with the same constant. (PC1) is trivial, but we have to modify (PC2), the triangle inequality.

One can prove that the modification is bounded, namely, for distinct X, Y, Z,

$$d^\pi_Y(X,Z) - 2\theta \le d_Y(X,Z) \le d^\pi_Y(X,Z).$$

It then follows from (PC2) that

(PC 2') $d_Y(X,Z) + d_Y(Z,W) + 4\theta \ge d_Y(X,W)$ (modified triangle inequality)

For a constant $K \ge \theta$ and distinct Y, Z we define a set, which is finite by (PC 4), as follows:

$$\mathbf{Y}_K(X,Z) = \{Y \in \mathbf{Y} | d_Y(X,Z) > K\}.$$

We say an element Y in $\mathbf{Y}_K(X,Z)$ is *between* X, Z. We are ready to define the projection complex.

Definition 2.1 (Projection complex). For a constant $K > 0$, the *projection complex* $\mathcal{P}_K(\mathbf{Y})$ is the following graph. The vertex set of $\mathcal{P}_K(\mathbf{Y})$ is $\mathbf{Y}$. Two distinct vertices X and Z are connected with an edge (of length 1) if $\mathbf{Y}_K(X,Z)$ is empty. Denote the distance function for this graph by $d(\ ,\ )$.

Note that for different values of K the spaces $\mathcal{P}_K(\mathbf{Y})$ are not necessarily quasi-isometric to each other (the vertex sets are the same, but for larger K there are more edges).

Here is the first main theorem.

Theorem 2.2. *Bestvina, Bromberg, and Fujiwara [2015, Theorem 3.16] Suppose functions $d^\pi_Y, Y \in \mathbf{Y}$ satisfy (PC1)- (PC4). Modify them to d_Y. If K is sufficiently large compared to θ, then the projection complex $\mathcal{P}_K(\mathbf{Y})$ is a quasi-tree.*

We make some comments on the proof. Suppose K is large. We first show that for X, Z, the subset $\mathbf{Y}_K(X,Z) \subset \mathbf{Y}$ gives a path between them in $\mathcal{P}_K(\mathbf{Y})$. In particular it implies

$$d(X,Z) \le |\mathbf{Y}_K(X,Z)| + 1.$$

In fact this path is a *quasi-geodesic* (ie, a quasi-isometric embedding of a segment) between the two points. This is the part where the modification to d_Y plays a role (without a modification the theorem is not true). Using the distances d_Y, we put a total order on $\mathbf{Y}_K(X,Z) \cup \{X,Z\}$ with X least and Z greatest: $Y < W$ if $d_W(X,Y) < \theta$. It takes some work to show this is well defined and gives a total order on the set $\mathbf{Y}_K(X,Z) \cup \{X,Z\}$. We then show this set form a path in $\mathcal{P}_K(\mathbf{Y})$ in this order. Recall the $\mathbb{H}^n$ example from the introduction. For axes X, Z, let σ be the shortest geodesic between them in $\mathbb{H}^n$. Then the set $\mathbf{Y}_K(X, Z)$ is roughly the collections of axes in $\mathbf{Y}$ that stay close to σ at least for distance K (cf. Figure 1). Because of (P 0), there is a bounded overlap between any two of them, so that there is an obvious order on $\mathbf{Y}_K(X, Z)$ in this picture.

There is also a lower bound of $d(X, Z)$. For a sufficiently large K' compared to K (roughly $5K$ is enough), if $Y \in \mathbf{Y}_{K'}(X, Z)$ then every geodesic from X to Z in $\mathcal{P}_K(\mathbf{Y})$ contains Y. This implies

$$d(X, Z) \geq |\mathbf{Y}_{K'}(X, Z)| + 1.$$

One can say that if X and Z has a "large projection" (ie, $\geq K'$) to Y, then every geodesic from X to Z has to pass Y.

2.2 Quasi-tree of metric spaces. To define a quasi-tree of metric spaces we need that each $Y \in \mathbf{Y}$ is a metric space. Here is the precise setup. Let $\mathbf{Y} = \{(Y, \rho_Y)\}$ be a collection of metric spaces and for each distinct $Y, Z \in \mathbf{Y}$ assume that we have sets $\pi_Y(Z) \subseteq Y$ and $\pi_Z(Y) \subseteq Z$. The π_Y are called *projection maps*. Fix a constant $\theta > 0$. Assume that for any $X \neq Y$,

(P 0) $\operatorname{diam}(\pi_X(Y)) \leq \theta$.

For any $X \neq Y \neq Z$, set

$$d_Y^\pi(X, Z) = \operatorname{diam}(\pi_Y(X) \cup \pi_Y(Z)),$$

where diam is ρ_Y-diameter. Then d_Y^π satisfy (PC1) and (PC2) trivially. Notice that (P0) implies (PC0). We assume that they also satisfy (PC3) and (PC4). Families of metric spaces with projection maps satisfying (PC0), (PC3) and (PC4) occur naturally in many contexts, see Examples 4.3.

Definition 2.3 (Quasi-tree of metric space). For $K > 0$, we build the *quasi-tree of metric spaces*, $\mathcal{C}_K(\mathbf{Y})$, by taking the union of the metric spaces in $\mathbf{Y}$ with an edge of length $L > 0$ (L depends on K) connecting every pair of points in $\pi_Y(X)$ and $\pi_X(Y)$ if $d_{\mathcal{P}_K(\mathbf{Y})}(X, Y) = 1$.

In other words, if $\mathbf{Y}_K(X, Y)$ is empty, we put edges between the sets $\pi_X(Y) \subset X$ and $\pi_Y(X) \subset Y$. $\mathcal{C}_K(\mathbf{Y})$ is a metric space. For example if each Y is a path metric space (which is the case in our applications), then $\mathcal{C}_K(\mathbf{Y})$ is path connected and there is an obvious metric on it. We denote the metric by $d_{\mathcal{C}_K(\mathbf{Y})}$.

To equip a group action on $\mathcal{C}_K(\mathbf{Y})$, consider a metric ρ (that is possibly infinite) on the disjoint union of elements of $\mathbf{Y}$ by setting $\rho(x_0, x_1) = \rho_X(x_0, x_1)$ if $x_0, x_1 \in X$, for some $X \in \mathbf{Y}$; and $\rho(x_0, x_1) = \infty$ if x_0 and x_1 are in different spaces in $\mathbf{Y}$. Assume that the group G acts isometrically on $\mathbf{Y}$ with this metric and that the projections π_X are G-invariant, i.e. $\pi_{gX}(gY) = g(\pi_X(Y))$. Then G naturally acts on $\mathcal{C}_K(\mathbf{Y})$ by isometries. This occurs in most examples.

2.3 Distance formula. In $\mathcal{P}_K(\mathbf{Y})$ we gave upper and lower bounds of $d(X, Z)$ using $\mathbf{Y}_K(X, Z)$ and $\mathbf{Y}_{K'}(X, Z)$. There are similar formula in $\mathcal{C}_K(\mathbf{Y})$ as follows. The projection $\pi_Y(X)$ is defined for $X \in \mathbf{Y}\backslash\{Y\}$, but since Y is a metric space in the current setting we extend the projection to each point $x \in X$ by $\pi_Y(x) = \pi_Y(X)$. Also, for a point $y \in Y$, we set $\pi_Y(y) = \{y\}$. The projection π_Y is now defined on $\mathcal{C}_K(\mathbf{Y})$, except for the edges.

Remark 2.4. Since $\mathcal{C}_K(\mathbf{Y})$ is a metric space we can also define $\pi_Y(x)$ using the nearest point projection. Then the Hausdorff distance between the sets we obtain in Y by the two different definitions is bounded. It is a part of Theorem 1.1.

As before, we then set $d_Y(x, z) = \operatorname{diam}(\pi_Y(x) \cup \pi_Y(z))$ for $x, z \in \mathcal{C}_K(\mathbf{Y})$. We also define

$$\mathbf{Y}_J(x, z) = \{Y \in \mathbf{Y} | d_Y(x, z) > J\}.$$

It is possible for X or Z, where $x \in X$ and $z \in Z$, to be in $\mathbf{Y}_J(x, z)$. Here is a *distance formula* in $\mathcal{C}_K(\mathbf{Y})$.

Theorem 2.5. *Bestvina, Bromberg, and Fujiwara [2015, Theorem 4.13] Let $K' > K$ be sufficiently large. Then for $x \in X, z \in Z$ we have*

$$\frac{1}{2} \sum_{W \in \mathbf{Y}_{K'}(x,z)} d_W(x, z) \le d_{\mathcal{C}_K(\mathbf{Y})}(x, z) \le 6K + 4 \sum_{W \in \mathbf{Y}_K(x,z)} d_W(x, z).$$

Some explanation is in order. In a way the lower bound is harder to get. Recall that if $W \in \mathbf{Y}_{K'}(X, Z)$ then every geodesic γ from X to Z in $\mathcal{P}_K(\mathbf{Y})$ has to visit the vertex W. In $\mathcal{C}_K(\mathbf{Y})$ the vertex W is replaced by the metric space W, so one changes γ to a path γ' in $\mathcal{C}_K(\mathbf{Y})$ by replacing the vertex W by a geodesic in W joining the subsets $\pi_W(x), \pi_W(z)$. Since the distance between those two sets is roughly $d_W(x, z)$, after those replacement the length of the path γ' is roughly bounded below by $\sum_{W \in \mathbf{Y}_{K'}(x,z)} d_W(x, z)$, which appears in the lower bound. The above formula is an analogy of the *Masur-Minsky distance formula* for a mapping class group (see Theorem 3.6).

3 Application to mapping class groups of surfaces

The family of mapping class groups of surfaces is an interesting object to study in Geometric group theory. We discuss applications of our construction to mapping class groups.

We recommend a book Farb and Margalit [2012] for a general reference of the theory of mapping class groups and a survey Minsky [2006] for the Masur-Minsky theory.

3.1 Definitions. Let $\Sigma = \Sigma_{g,b}$ be an orientable compact surface with genus g and b boundary components. The group of orientation preserving homeomorphisms of an oriented surface Σ to itself, taken modulo isotopy, is called the *mapping class group* of Σ, denoted by $MCG(\Sigma)$. $MCG(\Sigma)$ is a finitely presented group. An *essential* simple closed curve in Σ is an embedded circle in Σ that is homotopically non-trivial and not homotopic into the boundary (non-peripheral). We may just say simple closed curves. A mapping class, ie, an element in $MCG(\Sigma)$, that preserves a system of disjoint essential simple closed curves on Σ is called *reducible*. For example Dehn twists are reducible. Thurston classified the nontrivial conjugacy classes in $MCG(\Sigma)$ as reducible, finite-order, and *pseudo-Anosov*. A pseudo-Anosov mapping class does not preserve any finite set of closed curves, but instead preserves a pair of *measured geodesic laminations*.

For example, $MCG(\Sigma_{0,0})$ is trivial. For $\Sigma_{1,0}, \Sigma_{1,1}$, $MCG(\Sigma)$ is isomorphic to $SL(2,\mathbb{Z})$, and $MCG(\Sigma_{0,4})$ maps to $PSL(2,\mathbb{Z})$ with the kernel $(\mathbb{Z}/2\mathbb{Z})^2$. In particular they are hyperbolic groups. But $MCG(\Sigma)$ is not word-hyperbolic if $g \geq 2$ since it contains $\mathbb{Z}^2$ generated by commuting Dehn twists. One natural metric space $MCG(\Sigma)$ acts on by isometries is *the Teichmüller space*. The Teichmüller space for $\Sigma_{g,0}, g > 0$ is diffeomorphic to the Euclidean space of dimension $6g-6$, with the Teichmüller metric. The action by $MCG(\Sigma)$ is proper but not co-compact. There are a lot of negative curvature aspects on a Teichmuller space (cf. Minsky [1996]), but it is not δ-hyperbolic.

3.1.1 Curve graph. We recall another object which $MCG(\Sigma)$ acts on, and is useful for our approach. Let $\mathcal{C}_0(\Sigma)$ be the set of homotopy classes of essential simple closed curves and properly embedded simple arcs on Σ (when $\partial\Sigma$ is not empty) that are essential (not homotopic into $\partial\Sigma$). We then define the *curve graph*, $\mathcal{C}(\Sigma)$, to be the 1-complex obtained by attaching an edge to a pair of disjoint closed curves or arcs in $\mathcal{C}_0(\Sigma)$.

Remark 3.1. The graph we have constructed is often called the *curve and arc graph*, Masur and Minsky [2000]. The usual curve graph, whose vertices are only curves, is quasi-isometric to the curve and arc graph and so we will use the less cumbersome name of curve graph.

$\mathcal{C}(\Sigma)$ is connected, and $MCG(\Sigma)$ naturally acts on $\mathcal{C}(\Sigma)$ by automorphisms since homeomorphisms preserve disjointness, and the quotient is finite. The action is far from proper, but the homomorphism $MCG(\Sigma) \to Aut(\mathcal{C}(\Sigma))$ has at most finite kernel, and the index of the image is finite in $Aut(\mathcal{C}(\Sigma))$, (Ivanov, Luo, Korkmaz).

Masur and Minsky [1999], Masur and Minsky [2000] studies the geometry of a curve complex, and their work has a significant impact on the study of hyperbolic 3-manifolds and mapping class groups. The following result is the first important theorem (cf. Bestvina and Fujiwara [2007] for a non-orientable Σ). More recently, it is proved that δ is uniform for all Σ, Hensel, Przytycki, and Webb [2015].

Theorem 3.2. *$\mathcal{C}(\Sigma)$ is a δ-hyperbolic space, and $g \in MCG(\Sigma)$ is a hyperbolic isometry on $\mathcal{C}(\Sigma)$ if and only if g is a pseudo-Anosov element.*

Moreover, for a given surface Σ, there is a uniform positive lower bound on the translation length of a pseudo-Anosov element.

3.2 Applications. We explain a setting for us to apply projection complex to the study of $MCG(\Sigma)$. A set $\mathbf{Y}$ we take is very different from the $\mathbb{H}^n$ example in the introduction. $\mathbf{Y}$ is not a collection of subsets in some hyperbolic space, say, the curve graph of Σ, but it will be a certain collection of (isotopy classes of) *essential* subsurfaces $Y \subset \Sigma$. A subsurface is *essential* if it is π_1-injective and non-peripheral. To define a quasi-tree of metric spaces, we also need a metric space for each Y (here, we distinguish Y and the metric space associated to it). For that we take the curve complex $\mathcal{C}(Y)$ for each Y. We use *subsurface projection* to define projections between two subsurfaces in $\mathbf{Y}$, then apply our method after checking the axioms.

3.2.1 Subsurface projection. We say two essential subsurfaces *overlap* if $\partial Y \cap \partial Z \neq \emptyset$ (this means that the intersection is nonempty even after any isotopy). Following Masur and Minsky [2000], if Y and Z overlap, we define the *subsurface projection* $\pi_Y(Z) \subset \mathcal{C}(Y)$ by taking the intersection of ∂Z with Y and identifying homotopic curves and arcs. Also, when β is a simple closed curve that cannot be isotoped to be disjoint from Y, we similarly define a projection $\pi_Y(\beta) \subset \mathcal{C}(Y)$.

Also, we will need the curve graph for an essential simple closed curve γ. The definition has a somewhat different flavor and we do not recall a definition here. $\mathcal{C}(\gamma)$ is quasi-isometric to $\mathbb{Z}$, and the Dehn twist along γ, which leaves γ invariant, acts by a hyperbolic isometry. We will call γ a subsurface too. When a curve γ and the boundary of a subsurface Y intersect, we already defined $\pi_Y(\gamma)$ but we will also need $\pi_\gamma(Y) \subset \mathcal{C}(\gamma)$. More generally Y can be a curve. See for example Bestvina, Bromberg, and Fujiwara [2015, §5.1] for the precise definition of $\mathcal{C}(\gamma)$ and the projection.

Now we want to check that the projection complex axioms are satisfied.

Theorem 3.3. *Let $\mathbf{Y}$ be a collection of essential subsurfaces in Σ such that any two distinct subsurfaces intersect. For distinct $X, Y, Z \in \mathbf{Y}$ define $d^\pi_Y(X, Z) = \mathrm{diam}(\pi_Y(X) \cup \pi_Y(Z))$, where the diameter is measured in $\mathcal{C}(Y)$. Then $\{d^\pi_Y\}$ satisfy (P0)-(P2) for some constant $\theta(\Sigma)$, which depends only on Σ.*

Thus for every such family $\mathbf{Y}$ we obtain the projection complex $\mathcal{P}_K(\mathbf{Y})$ for a large K, and the quasi-tree of curve complexes $\mathcal{C}_K(\mathbf{Y})$ for $\{\mathcal{C}(Y), Y \in \mathbf{Y}\}$. We make comments on verifying the axioms (P0)-(P2) in this setting. Axiom (P0) follows easily from definitions. Axiom (P1) was established by J. A. Behrstock [2006]. We sometimes refer to Axiom (P1) in general as *Behrstock's inequality*. Axiom (P2) is by Masur-Minsky (a consequence of the Theorem 4.6 and Lemma 4.2 in Masur and Minsky [2000]). A central idea in Masur and Minsky [ibid.] is the notion of a *hierarchy* and this is used in the original verification of (P1) and (P2). This is a powerful tool but it is complicated to define and difficult to use. Leininger gave a very simple, hierarchy free proof of (P1) (see Mangahas [2010, 2013]) and also (P2) has a direct, hierarchy free proof (Bestvina, Bromberg, and Fujiwara [2015]).

3.2.2 Embedding MCG. Having Theorem 3.3, we now use Theorem 1.1 to embed the mapping class group in a *finite* product of *quasi-trees of curve complexes*. To use Theorem 3.3, we group the essential subsurfaces in Σ into finitely many subcollections $\mathbf{Y}^1, \mathbf{Y}^2, \cdots, \mathbf{Y}^k$, such that any X, Y in each family overlap, hence the projection $\pi_X(Y)$ is defined. The subcollections are the orbits of a certain subgroup $\mathcal{S}$ in $MCG(\Sigma)$, given in the following lemma.

Lemma 3.4 (Color preserving subgroup)**.** *Bestvina, Bromberg, and Fujiwara [ibid.]. There is a coloring $\phi : \mathcal{C}_0(\Sigma) \to F$ of the set of simple closed curves on Σ with a finite set F of colors so that if a, b span an edge then $\phi(a) \neq \phi(b)$. Moreover, there is a finite index subgroup $\mathcal{S}$ of the mapping class group $MCG(\Sigma)$ (where Σ is closed) such that every element of $\mathcal{S}$ preserves the colors.*

We call this subgroup *the color preserving subgroup*. Note that there are only finitely many $\mathcal{S}$-orbits of subsurfaces of Σ and any two subsurfaces in each $\mathcal{S}$-oribit overlap. Having Theorem 3.3, for each orbit $\mathbf{Y}^i$ we apply our construction to $\{\mathcal{C}(Y) | Y \in \mathbf{Y}^i\}$ and obtain $\mathcal{C}_K(\mathbf{Y}^i)$ for a large enough constant K. Everything (for example projection $\pi_X(Y)$) is done equivariantly in the construction, so that we have an equivariant orbit map

$$\Phi : MCG(\Sigma) \to \mathcal{C}_K(\mathbf{Y}^1) \times \mathcal{C}_K(\mathbf{Y}^2) \times \cdots \times \mathcal{C}_K(\mathbf{Y}^k),$$

sending $g \in MCG(\Sigma)$ to $g(o)$, where o is an arbitrary base point in the product. Note that an element in $\mathcal{S}$ preserves each factor, while other elements permute the factors. The choice of a base point will not be important for our purpose. We put the ℓ^1-metric on the product.

Since $MCG(\Sigma)$ is finitely generated, the map Φ is a Lipschitz map. Moreover, by some compactness argument regarding the set of curves on Σ, one can show Φ is a coarse embedding. A map between two metric spaces $f : X \to Y$ is a *coarse embedding* if there are constants A, B and a function $\Phi : [0, \infty) \to [0, \infty)$ with $\Phi(t) \to \infty$ as $t \to \infty$ such that

$$\Phi(|x - x'|) \leq |f(x) - f(x')| \leq A\,|x - x'| + B.$$

Moreover, it turns out that Φ is a quasi-isometric (QI) embedding. As each factor is hyperbolic by Theorem 1.2 (ii), we have the following theorem:

Theorem 3.5. *Bestvina, Bromberg, and Fujiwara [ibid.]. $MCG(\Sigma)$ equivariantly quasi-isometrically embeds in a finite product of hyperbolic spaces.*

The argument to show that Φ is a QI-embedding is by reinterpreting the remarkable Masur-Minsky distance formula (Theorem 6.12 in Masur and Minsky [2000]). To state it, let α be a finite *binding* collection of simple closed curves in Σ, ie, every essential curve in Σ intersects at least one curve in α. For x, M, the number $[x]_M$ is defined as x if $x > M$ and as 0 if $x \leq M$. Fix a finite generating set of $MCG(\Sigma)$, and let $|g|$ be the word norm.

Theorem 3.6 (Masur-Minsky distance formula)**.** *Suppose M is sufficiently large. Then there exist K, L such that for any $g \in MCG(\Sigma)$,*

$$\frac{1}{K}|g| - L \leq \sum_{Y} [d_{\mathcal{C}(Y)}(\pi_Y(\alpha), \pi_Y(g(\alpha)))]_M \leq K|g| + L,$$

where the sum is over all essential subsurfaces Y in Σ.

After arranging $K' = M$, notice that the sum in the middle of the above theorem appears in the left hand side (to be precise, we add them over all $\mathbf{Y}^i$'s) of the distance formula in a quasi-tree of metric spaces (Theorem 2.5) with $x = \pi_Y(\alpha), z = \pi_Y(g(\alpha))$. Combing those two estimates we obtain a desired estimate to show that Φ is a QI-embedding.

The following result follows easily from the definition of *asymptotic cones* (see J. Behrstock, Druţu, and Sapir [2011b,a]) and Theorem 3.5 since the asymptotic cone of a hyperbolic space is an $\mathbb{R}$-tree.

Theorem 3.7 (Behrstock-Druţu-Sapir)**.** *Every asymptotic cone of $MCG(\Sigma)$ embeds by a bi-Lipschitz map in a finite product of $\mathbb{R}$-trees.*

In fact they prove more including some information on the geometry of the image of the embedding, but their theorem does not imply Theorem 3.5. They use the notion of *tree-graded space* introduced in Druţu and Sapir [2005].

3.2.3 Asymptotic dimension. We discuss a further application to asymptotic dimension, which we stated as Theorem 1.4. We recall a few basic properties. Let $\mathcal{X}, \mathcal{Y}$ be metric spaces. If $\mathcal{X} \subset \mathcal{Y}$, with a metric on $\mathcal{Y}$ restricted to $\mathcal{X}$, then $\operatorname{asdim}(\mathcal{X}) \leq \operatorname{asdim}(\mathcal{Y})$.

We have Product Formula: $\operatorname{asdim}(\mathcal{X} \times \mathcal{Y}) \leq \operatorname{asdim}(\mathcal{X}) + \operatorname{asdim}(\mathcal{Y})$. It is straightforward from the definition that the asymptotic dimension is not only a quasi-isometric invariant but is also a *coarse invariant*. In particular $\operatorname{asdim}(\mathcal{X}) \leq \operatorname{asdim}(\mathcal{Y})$ if there exists a coarse embedding $f : \mathcal{X} \to \mathcal{Y}$ (Roe [2003]).

It is a theorem of Bell and Fujiwara [2008] that each curve complex has finite asymptotic dimension. Thus from Theorems 1.2 (iii) and 3.5 we obtain the following theorem, which motivated the work Bestvina, Bromberg, and Fujiwara [2015]:

Theorem 3.8. *(Theorem 1.4) Let Σ be a closed orientable surface, possibly with punctures. Then* $\operatorname{asdim}(MCG(\Sigma)) < \infty$.

The exact value of $\operatorname{asdim}(MCG(\Sigma))$ is unknown. Webb Webb [2015] found explicit bounds on the asymptotic dimension of curve complexes, which was improved to a linear bound by Bestvina and Bromberg [n.d.] by a different method. As a consequence, $\operatorname{asdim}(MCG(\Sigma))$ is bounded by an exponential function in the *complexity* of the surface, $\kappa(\Sigma_{g,b}) = 3g + b$.

We can also prove:

Theorem 3.9. *Bestvina, Bromberg, and Fujiwara [2015]. The Teichmüller space of Σ, with either the Teichmüller metric or the Weil-Petersson metric, has finite asymptotic dimension.*

3.2.4 Dehn twists as hyperbolic isometry. When X is a quasi-tree, an isometry with unbounded orbits is always hyperbolic, Manning [2006]. The following theorem uses the observation that the $MCG(\Sigma)$-orbit of a curve α in a surface Σ of even genus that separates Σ into subsurfaces of equal genus consists of pairwise intersecting curves. We then form a quasi-tree of metric spaces for the collection of those curves, viewed as subsurfaces as we explained (and their curve complexes), where the Dehn twist along α will be hyperbolic with an axis $\mathcal{C}(\alpha)$.

Theorem 3.10. *Bestvina, Bromberg, and Fujiwara [2015]. The mapping class groups in even genus can act on quasi-trees with at least one Dehn twist having unbounded orbits.*

In the case of odd genus one has to pass to the color preserving subgroup $\mathcal{S}$. In any case, it follows that for each Dehn twist g, the word norm $|g^n|$ has linear growth on n in $MCG(\Sigma)$ (not only in $\mathcal{S}$). Thus we recover the theorem by Farb, Lubotzky, and Minsky [2001] (the other types of elements are easy):

Theorem 3.11. *For each element g of infinite order in $MCG(\Sigma)$, the word norm $|g^n|$ has linear growth.*

3.2.5 Promoting actions to CAT(0) spaces or trees. Theorem 3.10 provides a sharp contrast to a result of Bridson [2010], who showed that in semi-simple actions of mapping class groups (of genus >2) on complete CAT(0) spaces Dehn twists are always elliptic. A group action is *semi-simple* if each element has either a bounded orbit or positive translation length.

By a *thickening* of a metric space X we mean a quasi-isometric embedding $X \to Y$. When X is a graph with edges of length 1 and $d \geq 1$, there is a particular thickening $X \to P_d(X)$ called the *Rips complex* of X. The space $P_d(X)$ is a simplicial complex with the same vertex set as X and with simplices consisting of finite collections of vertices with pairwise distance at most d. The Dehn twists that are hyperbolic in Theorem 3.10 stay hyperbolic in any thickning of the quasi-tree. Now by the theorem of Bridson, a thickning is never CAT(0). This give the following theorem:

Theorem 3.12. *Bestvina, Bromberg, and Fujiwara [2015]. There is an isometric action of a group on a graph X which is a quasi-tree such that no equivariant thickening admits an equivariant $CAT(0)$ metric. In particular, for no $d \geq 1$, the Rips complex $P_d(X)$ admits an equivariant $CAT(0)$ metric.*

We make some comments on the background. It is a long-standing open question whether every hyperbolic group acts co-compactly and properly by isometries on a CAT(0) space. One approach is to consider the Rips complex $P_d(X)$ for the Cayley graph X of the group and large d. Theorem 3.12 is not a counterexample to this approach since our X is not locally finite, but it does point out difficulties. Note that in light of Mosher, Sageev, and Whyte [2003] the quasi-trees that arise in our construction are necessarily locally infinite, since otherwise we would be able to promote our group actions on quasi-trees to group actions on simplicial trees without fixed points, which is not possible for certain groups, for example, non-elementary hyperbolic groups that have property (T), eg, uniform lattices in $\mathrm{Sp}(n,1)$.

3.2.6 Uniform uniform exponential growth of $MCG(\Sigma)$. We discuss applications to the exponential growth of groups. Let Γ be a group and S a finite set in Γ. Assume that $1 \in S$ and $S = S^{-1}$. Set

$$h(S) := \lim_{n\to\infty} \frac{1}{n} \log |S^n|.$$

Let Γ be a finitely generated group. Set $h(\Gamma) = \inf_S \{h(S) | \langle S\rangle = \Gamma\}$, where S runs over all finite generating subsets. If $h(\Gamma) > 0$ we say Γ has *uniform exponential growth*, of *growth rate* $h(\Gamma)$.

Using quasi-trees of metric spaces, Breuillard and Fujiwara [n.d.] recovers the following theorem by Mangahas, Mangahas [2010]:

Theorem 3.13 (UUEG of MCG). *Let Σ be a compact oriented surface possibly with punctures. Then there exists a constant $N(\Sigma)$ such that for any finite set $S \subset MCG(\Sigma)$ with $S = S^{-1}$, either $\langle S \rangle$ is virtually abelian or S^N contains g, h that produces a free semi-group. In particular: $h(S) \geq \frac{1}{N} \log 2$.*

It follows that for each surface Σ there is a constant $c(\Sigma) > 0$ such that for any finite set S in $MCG(\Sigma)$, $h(S) \geq c(\Sigma)$ unless $\langle S \rangle$ is virtually abelian. We say $MCG(\Sigma)$ has "uniform uniform exponential growth" (UUEG). The proof of Theorem 3.13 in Breuillard and Fujiwara [n.d.] applies a standard "Ping-Pong" argument to the actions of the color preserving subgroup $\mathcal{S}$ on the $\mathcal{C}_K(\mathbf{Y}^i)$'s. A key property of those actions is that any non-trivial element of infinite order in $\mathcal{S}$ is hyperbolic for at least one of the actions. Mangahas' proof is different and does not use our complexes. Also she proves more, that g, h produces a free group of rank-two, maybe for a larger constant $N(\Sigma)$.

3.2.7 Stable commutator length. Let G be a group, and $[G, G]$ its commutator subgroup. For an element $g \in [G, G]$, let $cl(g) = cl_G(g)$ denote the *commutator length* of g, the least number of commutators whose product is equal to g. We define $cl(g) = \infty$ for an element g not in $[G, G]$. For $g \in G$, the *stable commutator length*, $\mathrm{scl}(g) = \mathrm{scl}_G(g)$, is defined by

$$\mathrm{scl}(g) = \liminf_{n \to \infty} \frac{cl(g^n)}{n} \leq \infty.$$

It is clear that $\mathrm{scl}(g^n) = n\,\mathrm{scl}(g)$ and $\mathrm{scl}(hgh^{-1}) = \mathrm{scl}(g)$. We recommend a monograph Calegari [2009] as a reference on scl.

One theme in the subject is to classify elements g in a given group for which $scl(g) > 0$. To verify $scl(g) > 0$ for an element g, the notion of quasi-morphisms is useful. A function $H : G \to \mathbb{R}$ is a *quasi-morphism* if

$$\Delta(H) := \sup_{x,y \in G} |H(xy) - H(x) - H(y)| < \infty.$$

$\Delta(H)$ is called the *defect* of H. It is a simple exercise to show that if there exists a quasi-morphism $f : G \to \mathbb{R}$ which is unbounded on the powers of g, then $scl(g) > 0$. Also, the converse holds by so called *Bavard duality*. For example, Brooks [1981] showed that in free groups G, $scl(g) > 0$ for every nontrivial element g by constructing a quasi-morphism $f : G \to \mathbb{R}$ which is unbounded on the powers of g.

On the other hand, it is straightforward from the definition that in the following situations $cl(g^n)$ is bounded and therefore $scl(g) = 0$:

(a) g has finite order,

(b) more generally, g is *achiral*, i.e. g^k is conjugate to g^{-k} for some $k \neq 0$.

For example, Epstein and Fujiwara [1997], generalizing the Brooks construction of quasi-morphisms, proved that in hyperbolic groups G the above obstruction (b) is the only one, namely, if g is *chiral* (i.e. not achiral) then $scl(g) > 0$. They use the hyperbolicity of the Cayley graph of G to construct a suitable quasi-morphism for g.

There are three more conditions one can directly check from the definition for $cl(g^n)$ being bounded and therefore $scl(g) = 0$, cf. Bestvina, Bromberg, and Fujiwara [2016b]:

(c) $g = g_1 g_2^{-1}$ such that $g_1 g_2 = g_2 g_1$, and g_1 is conjugate to g_2,

(d) more generally, g is expressed as a commuting product $g = g_1 \cdots g_p$ such that $g_i^{n_i}$ are all conjugate for some $n_i \neq 0$ and that $\sum_i \frac{1}{n_i} = 0$,

(e) $g = g_1 \cdots g_p$ is a commuting product and $cl(g_i^n)$ are bounded for all i.

Now, we are interested in the question to decide when $scl(g) = 0$ for $g \in MCG(\Sigma)$. Some partial answers were known. Using 4-manifold invariants, Endo and Kotschick [2001] and Korkmaz [2004] prove that $scl(g) > 0$ if g is a Dehn twist. Endo and Kotschick [2007] also note the obstruction (c): in $MCG(\Sigma)$ for example this occurs if g_1, g_2 are Dehn twists in disjoint curves in the same $MCG(\Sigma)$-orbit. By contrast, Calegari and Fujiwara [2010] prove that if g is pseudo-Anosov and chiral then $scl(g) > 0$, ie, (b) is the only obstruction among pseudo-Anosov elements. They use Theorem 3.2 to construct a suitable quasi-morphism for g.

We state our result in the following vague form. See the precise statement in Bestvina, Bromberg, and Fujiwara [2016b]. It covers all cases in a unified way and in particular we recover the result on Dehn-twists.

Theorem 3.14. *Let $G < MCG(\Sigma)$ be a subgroup of finite index and $g \in G$. Then there is a characterization of elements g with $scl_G(g) > 0$ in terms of the "Nielsen-Thurston form" of g.*

The new and more complicated case is on a reducible element g. By Nielsen-Thurston form, a power of such g is written as a commuting product of powers of Dehn twists and pseudo-Anosov maps on disjoint subsurfaces on Σ, after removing a system of g-invariant curves. We argue that either $cl(g^n)$ is bounded applying (a)-(e) to the Nielsen-Thurston form of g, or else we can construct a suitable quasi-morphism on $\mathcal{S} \cap G$ to show $scl(g) > 0$. A key point is g is hyperbolic on one of the $\mathcal{C}_K(\mathbf{Y}^i)$'s. Also we prove the following theorem. The *Torelli subgroup* is the kernel of the action of $MCG(\Sigma)$ on $H_1(\Sigma, \mathbb{Z})$. It has infinite index as a subgroup.

Theorem 3.15. *If $G = \mathcal{T}$ is the Torelli subgroup in $MCG(\Sigma)$ and $1 \neq g \in G$ then $scl_G(g) > 0$.*

4 Other applications

4.1 Contracting geodesics and WPD. In the introduction we explained how we build a quasi-tree from a discrete group G of isometries of $\mathbb{H}^n$, and extracted the axioms we need. The axioms are satisfied since $\mathbb{H}^n$ is hyperbolic and G is discrete. We can relax the assumptions keeping the axioms satisfied.

Let X be a geodesic space, and $Y \subset X$ a subset. For a constant $B > 0$ we say that Y is *B-contracting* if the nearest point projection (this is a *coarse map*, ie, the image of a point may contain more than one point) to Y of any metric ball disjoint from Y has diameter bounded by B. See Bestvina and Fujiwara [2009]. For example if X is δ-hyperbolic and Y is a geodesic, then Y is 10δ-contracting. Let γ be a hyperbolic isometry of X, and let O be the γ-orbit of a point x in X. We say γ is *rank 1* if O is B-contracting for some B. This notion does not depend on the choice of x, and also, one can take an axis of γ, if it exists, instead of an orbit O for an equivalent definition, Bestvina, Bromberg, and Fujiwara [n.d.]. Any hyperbolic isometry on a hyperbolic space is rank 1.

Assume that a group Γ acts by isometries on a geodesic metric space X, and $\gamma \in \Gamma$ acts hyperbolically. We say γ is a *WPD element* if for all $D > 0$ and $x \in X$ there exists $M > 0$ such that the set

$$\{g \in \Gamma \mid d(x, g(x)) \le D, d(f^M(x), gf^M(x)) \le D\}$$

is finite. Bestvina and the author Bestvina and Fujiwara [2002] introduced this notion and proved that every pseudo-Anosov element in $MCG(\Sigma)$ is WPD on $\mathcal{C}(\Sigma)$, in view of an application to computing the *second bounded cohomology* of a subgroup A, $H_b^2(A, \mathbb{R})$, in $MCG(\Sigma)$. Note that if the action of G on X is proper, then any hyperbolic isometry is WPD. WPD stands for *weak proper discontinuous*. There is even a weaker notion, called *WWPD*, introduced in Bestvina, Bromberg, and Fujiwara [2015]. This notion is needed to prove Theorem 3.14. Delzant [2016] found an application of WWPD and the projection complex to Kähler groups. Handel and Mosher [n.d.] found an application of WWPD to computing the second bounded cohomology of subgroups in $Out(F_n)$.

The $\mathbb{H}^n$ example in the introduction is a special case of the following theorem.

Theorem 4.1. *Bestvina, Bromberg, and Fujiwara [2015], Bestvina, Bromberg, and Fujiwara [n.d.]. Let Γ act on a geodesic metric space X such that some $\gamma \in \Gamma$ is a hyperbolic WPD element. Assume the γ-orbit of some point, O, is B-contracting for some B (ie, γ is rank 1). Then the collection of parallel classes of Γ-translates of the orbit O with nearest point projections satisfies (P0)-(P2) and thus Γ acts on a quasi-tree, Q. In addition, in this action γ is a hyperbolic WPD element.*

Some explanation is in order. We say that two orbits are *parallel* if their Hausdorff distance is finite. The quasi-tree Q in the theorem is a quasi-tree of metric spaces for the

Γ-translates of O, each of which is quasi-isometric to a line. We do not assume that X is hyperbolic nor CAT(0). The main part of the proof consists of verifying (P0)-(P2) and applying Theorem 1.1 in this situation.

Remark 4.2. If an infinite geodesic γ bounds a Euclidean half plane, then clearly γ is not B-contracting for any B. This condition, *the flat half plane condition*, was important in the study of $CAT(0)$-space X, see Ballmann [1995]. If X is a locally compact $CAT(0)$-space and γ is an axis of a hyperbolic isometry g then γ being B-contracting for some B is equivalent to that γ does not bound a flat half plane, Bestvina and Fujiwara [2009]. See also Charney and Sultan [2015] for a recent development. The flat half plane conditions and the notion of rank 1 are first defined in the study of Riemannian manifolds of nonpositive sectional curvature.

Examples 4.3. The following examples all satisfy Theorem 4.1. One considers the translates of an axis, or more generally the orbit of a point, of a hyperbolic WPD element, γ.

(1) As a generalization of discrete subgroups in $\mathbb{H}^n$, Γ is a group of isometries of a δ-hyperbolic space X that contains a hyperbolic, WPD element, γ. For example, in the action of a hyperbolic group on its Cayley graph, any element of infinite order is hyperbolic and WPD. The class of hyperbolic groups contains many groups with Kazhdan's property (T) and therefore every isometric action on a simplicial tree has a fixed point (cf. de la Harpe and Valette [1989]). For the action of $MCG(\Sigma)$ on the curve complex, every pseudo-Anosov element is hyperbolic and WPD Bestvina and Fujiwara [2002].

(2) Let G be the fundamental group of a *rank-1 manifold* M, ie, M is a complete Riemannian manifold of non-positive curvature of finite volume such that the universal cover X of M is not a Riemannian product nor a symmetric space of non-compact type of rank at least two. Then G properly acts on X and by the *Rank rigidity theorem* Ballmann [1985], G contains a hyperbolic isometry γ which is rank-1.

(3) Γ is a discrete group of isometries (i.e. the group action is proper) of a $CAT(0)$-space that contains a hyperbolic rank-1 element γ. Also, pseudo-Anosov mapping classes are rank 1 elements in the action of $MCG(\Sigma)$ on the Weil-Petersson completion of Teichmüller space, which is CAT(0). Those elements are WPD although the action of $MCG(\Sigma)$ is not properly discontinuous. See Bestvina and Fujiwara [2009]. There are classifications of rank 1 elements in Coxeter groups Caprace and Fujiwara [2010], right angled Artin groups J. Behrstock and Charney [2012] and cube complexes Caprace and Sageev [2011].

(4) Γ is $MCG(\Sigma)$ acting on Teichmüller space with Teichmüller metric, and γ is a pseudo-Anosov mapping class. By Minsky [1996] the axis of γ is B-contracting. It is WPD since the action is properly discontinuous.

(5) $\Gamma = Out(F_n)$ acting on Culler-Vogtmann's Outer space CV_n Culler and Vogtmann [1986], equipped with the Lipschitz metric (not symmetric, see Algom-Kfir and Bestvina [2012]). The action is properly discontinuous. CV_n is not δ-hyperbolic. See for example Vogtmann [2006] for more information on $Out(F_n)$ and Outer space. As an analogue of a pseudo-Anosov element in a mapping class group, an element f of $Out(F_n)$ is *fully irreducible* if there are no conjugacy classes of proper free factors of F_n which are f-periodic. Such elements are hyperbolic with axes in CV_n, see Bestvina [2011], which are B-contracting (Algom-Kfir [2011]) .

(6) *The Cremona group*, G, of all birational transformations of the projective plane $\mathbb{P}^2_k$, where k is an algebraically closed field. G acts on a hyperbolic space, and it contains a hyperbolic WPD element, which was shown by Cantat and Lamy [2013]. It then follows that Cremona groups are not simple. We explain this implication in the next section.

4.2 Acylindrically hyperbolic groups and hyperbolically embedded subgroups. Dahmani, Guirardel, and Osin [2017] introduced the notion of *hyperbolically embedded subgroups*, a generalization of the concept of a *relatively hyperbolic group* (see their paper for the precise definition). They proved

Theorem 4.4. *If G is not virtually cyclic and acts on a hyperbolic space X such that G contains γ that is hyperbolic and WPD, then G contains a proper infinite hyperbolically embedded subgroup H.*

Here, we can take H to be virtually cyclic containing $\langle\gamma\rangle$. They use projection complex as a key tool in the argument. They further proved that for a sufficiently large N, γ^N normally generates a free subgroup (of maybe infinite rank) whose non-trivial elements are all hyperbolic on X. In particular G is not simple. This is the implication we mentioned in Examples 4.3 (6), and it applies to groups G in Examples 4.3 by Theorem 4.1. Also, by this method, they produce a free normal subgroup in $MCG(\Sigma)$, unless it is virtually cyclic, whose non-trivial elements are all pseudo-Anosov.

An isometric group action is *acylindrical* if for every $D > 0$ there exist $R, N > 0$ such that $d(x, y) \geq R$ implies that the set

$$\{g \in G \mid d(x, g(x)) \leq D, d(y, g(y)) \leq D\}$$

has cardinality at most N. Notice that this property implies WPD for any hyperbolic element. If a group action is proper and co-compact then it is acylindrical. Sela [1997] introduced the acylindricity of a group action on simplicial trees, then Bowditch [2008] formulated this definition for hyperbolic spaces and proved that the action of $MCG(\Sigma)$ on $\mathcal{C}(\Sigma)$ is acylindrical. Based on this definition, Osin [2016] develops a theory of *acylindrically hyperbolic groups*: these are groups that admit a non-elementary acylindrical isometric action on a hyperbolic space. Here, an action of G on a hyperbolic spce X is *non-elementary* if the limit set of the G-orbit of a point in X contains at least three points. He proved the following theorem.

Theorem 4.5. *Osin [ibid.] Let a group Γ, which is not virtually cyclic, act on a δ-hyperbolic metric space X such that $\gamma \in \Gamma$ is a hyperbolic WPD element. Then Γ is an acylindrically hyperbolic group. Thus all groups in Examples 4.3 are acylindrically hyperbolic.*

We make comments on his argument. By Theorem 4.4, G contains a hyperbolically embedded subgroup H. To construct a hyperbolic space for G to act on acylindrically, he uses an idea similar to projection complex with $\mathbf{Y}$ to be translates of an orbit of H. It has been improved in Balasubramanya [n.d.] so that the hyperbolic space in Theorem 4.5 can be taken to be a quasi-tree. In Bestvina, Bromberg, Fujiwara, and Sisto [n.d.] we recover this improvement by a different axiomatic construction: we start with $\mathbf{Y}$, the translates of an orbit of γ in X, which satisfies (PC0) -(PC4), then slightly change the definition of the projection $\pi_X(Y)$. For this new projection, the resulting projection complex by the usual definition is a quasi-tree, acted by G acylindrically.

4.3 Actions on CAT(0) square complex. Recall that Burger and Mozes [2000] constructed an example of a simple group, which acts freely and co-compactly on the product of two trees. Thus the quotient is a finite non-positively curved square complex with finitely-presented, infinite simple fundamental group. A square complex Z, built from unit Euclidean squares, is *non-positively curved* if the universal cover is CAT(0). Caprace and Delzant pointed out the following curious corollary of Theorem 4.1, which can be thought of a converse of the Burger-Mozes theorem.

Corollary 4.6 (see Bestvina, Bromberg, and Fujiwara [2015]). *Suppose Z is a finite non-positively curved square complex with no free edges whose fundamental group is simple. Then the universal cover $\tilde{Z}$ is isometric to the product of two trees.*

They argue that by the Ballmann-Brin Rank Rigidity Theorem Ballmann and Brin [1995, Th C] (see also Caprace and Sageev [2011]) the universal cover $\tilde{Z}$ is either the product of two trees or the deck group contains a rank 1 element. But in the latter case, by Theorem 4.1, $\pi_1(Z)$ acts on a quasi-tree and contains a hyperbolic WPD element γ. Also, $\pi_1(Z)$ is non virtually cyclic since it is simple and torsion-free. Now as we explained $\pi_1(Z)$ is not simple, impossible.

4.4 Bounded cohomology and QFA. Manning [2005] gave a construction of an action of a group G on a quasi-tree starting with a quasi-morphism $G \to \mathbb{R}$ but it is not clear when such actions are *non-elementary* (i.e. have unbounded orbits and do not fix an end nor a pair of ends). Groups G in Examples 4.3 have isometric actions on quasi-trees, and if G is non-elementary (ie, not virtually cyclic), then the action is non elementary. Conversely, if one has actions of a group G on a quasi-tree (with a hyperbolic WPD element), one can use such actions to give unified constructions of quasi-morphisms on G (cf. Epstein and Fujiwara [1997], Fujiwara [2000], Fujiwara [1998], Bestvina and Fujiwara [2002]), and even *quasi-cocycles* with coefficients in unitary representations in "uniformly convex" Banach spaces, for example, the regular representation on $\ell^2(G)$, which is of particular importance (see Monod [2006]). As a consequence, we prove

Theorem 4.7. *Bestvina, Bromberg, and Fujiwara [2016a] Let G be an acylindrically hyperbolic group with no non-trivial finite normal subgroup, and ρ a unitary representation of G in a uniformly convex Banach space, then the second bounded cohomology $H^2_b(G;\rho)$ is infinite dimensional.*

By contrast, there are many groups that do not admit nontrivial (namely, orbits are unbounded) actions on a quasi-tree. A group G satisfies *QFA* if every action on a quasi-tree has bounded orbits. For example, $SL_n(\mathbb{Z})$, $n \geq 3$ satisfies QFA, Manning [2006]. More recently, Haettel [n.d.] proves that if G is a lattice in (a product of) a higher rank semi-simple Lie group with finite center, then G satisfies QFA. He even proved that an action on any hyperbolic space X by such G has either a bounded orbit in X or has a fixed point in the ideal boundary of X.

4.5 $Out(F_n)$. A version of Theorem 3.5 for $Out(F_n)$ is known. There are several analogs of the curve graph $\mathcal{C}(\Sigma)$, for example the complex of free factors and the complex of free splittings. Both have been shown to be hyperbolic, the former in Bestvina and Feighn [2014a] and the latter in Handel and Mosher [2013]. The analog of subsurface projections was defined by Bestvina-Feighn in Bestvina and Feighn [2014b] and they show using the projection complex technique:

Theorem 4.8. *Bestvina and Feighn [ibid.] $Out(F_n)$ acts isometrically on a finite product of hyperbolic spaces so that every element of "exponential growth" acts with positive translation length.*

It is unknown if $Out(F_n)$ acts on a finite product of hyperbolic spaces that gives a QI-embedding. While a finite product of hyperbolic spaces satisfies a quadratic *isoperimetric inequality*, it is known that the isoperimetric inequality of $Out(F_n)$ is exponential (cf. Bridson and Vogtmann [1995]), but that is not an obstruction for QI-embeddings because we do not require that an embedding is (quasi-)convex. Theorem 1.4 (finiteness of asymptotic dimension) is unknown for $Out(F_n)$, but recently Bestvina-Guirardel-Horbez proved that $Out(F_n)$ is *boundary amenable*, therefore satisfies the Novikov conjecture on higher signatures Bestvina, Guirardel, and Horbez [n.d.].

4.6 Farrell-Jones conjecture for MCG. Bartels and Bestvina [n.d.] prove the *Farrell-Jones Conjecture* for mapping class groups:

Theorem 4.9 (Bestvina-Bartels)**.** *The mapping class group $Mod(\Sigma)$ of any oriented surface Σ of finite type satisfies the Farrell-Jones Conjecture.*

The main step of the proof is the verification of a regularity condition, called *finite $\mathfrak{F}$-amenability* (see Bartels and Bestvina [ibid.] for a precise definition). Using subsurface projections by Masur-Minsky, combined with the projection complex technique, they prove the action of $MCG(\Sigma)$ on the space $\mathfrak{PMF}$ of projective measured foliations on Σ is finitely $\mathfrak{F}$-amenable, for a certain family $\mathfrak{F}$ of subgroups in $MCG(\Sigma)$. Theorem 4.9 is then a consequence of the axiomatic results of Lück, Reich and Bartels for the Farrell-Jones Conjecture (cf. Bartels and Lück [2012]) and an induction on the complexity of the surface.

$MCG(\Sigma)$ has finite asymptotic dimension (Theorem 1.4). As we said this implies the integral Novikov conjecture, i.e., the integral injectivity of the assembly maps in algebraic K-theory and L-theory relative to the family of finite subgroups. This is related to the Farrell-Jones conjecture (see Bartels and Bestvina [n.d.]).

References

Yael Algom-Kfir (2011). “Strongly Contracting Geodesics in Outer Space”. *Geometry and Topology* 15.4, pp. 2181–2234 (cit. on p. 1126).

Yael Algom-Kfir and Mladen Bestvina (2012). “Asymmetry of outer space”. *Geom. Dedicata* 156, pp. 81–92. MR: 2863547 (cit. on p. 1126).

Sahana Balasubramanya (n.d.). “Acylindrical group actions on quasi-trees”. arXiv: 1602.03941 (cit. on p. 1127).

Werner Ballmann (1985). “Nonpositively curved manifolds of higher rank”. *Ann. of Math. (2)* 122.3, pp. 597–609. MR: 819559 (cit. on p. 1125).

– (1995). *Lectures on spaces of nonpositive curvature*. Vol. 25. DMV Seminar. With an appendix by Misha Brin. Birkhäuser Verlag, Basel, pp. viii+112. MR: 1377265 (cit. on pp. 1111, 1125).

Werner Ballmann and Michael Brin (1995). “Orbihedra of nonpositive curvature”. *Inst. Hautes Études Sci. Publ. Math.* 82, 169–209 (1996). MR: 1383216 (cit. on p. 1127).

Arthur Bartels and Mladen Bestvina (n.d.). “The Farrell-Jones Conjecture for mapping class groups”. arXiv: 1606.02844 (cit. on p. 1129).

Arthur Bartels and Wolfgang Lück (2012). "The Borel conjecture for hyperbolic and CAT(0)-groups". *Ann. of Math. (2)* 175.2, pp. 631–689. MR: 2993750 (cit. on p. 1129).

Jason A. Behrstock (2006). "Asymptotic geometry of the mapping class group and Teichmüller space". *Geom. Topol.* 10, pp. 1523–1578. MR: 2255505 (cit. on p. 1118).

Jason A. Behrstock and Yair N. Minsky (2008). "Dimension and rank for mapping class groups". *Ann. of Math. (2)* 167.3, pp. 1055–1077. MR: 2415393 (cit. on p. 1110).

Jason Behrstock and Ruth Charney (2012). "Divergence and quasimorphisms of right-angled Artin groups". *Math. Ann.* 352.2, pp. 339–356. MR: 2874959 (cit. on p. 1125).

Jason Behrstock, Cornelia Druţu, and Mark Sapir (2011a). "Addendum: Median structures on asymptotic cones and homomorphisms into mapping class groups". *Proc. Lond. Math. Soc. (3)* 102.3, pp. 555–562. MR: 2783136 (cit. on p. 1119).

– (2011b). "Median structures on asymptotic cones and homomorphisms into mapping class groups". *Proc. Lond. Math. Soc. (3)* 102.3, pp. 503–554. MR: 2783135 (cit. on p. 1119).

Gregory C. Bell and Koji Fujiwara (2008). "The asymptotic dimension of a curve graph is finite". *J. Lond. Math. Soc. (2)* 77.1, pp. 33–50. MR: 2389915 (cit. on p. 1120).

Mladen Bestvina (2011). "A Bers-like proof of the existence of train tracks for free group automorphisms". *Fund. Math.* 214.1, pp. 1–12. MR: 2845630 (cit. on p. 1126).

Mladen Bestvina, Ken Bromberg, and Koji Fujiwara (2015). "Constructing group actions on quasi-trees and applications to mapping class groups". *Publ. Math. Inst. Hautes Études Sci.* 122, pp. 1–64. MR: 3415065 (cit. on pp. 1105, 1107–1111, 1113, 1115, 1117–1121, 1124, 1127).

– (2016a). "Bounded cohomology with coefficients in uniformly convex Banach spaces". *Comment. Math. Helv.* 91.2, pp. 203–218. MR: 3493369 (cit. on p. 1128).

– (2016b). "Stable commutator length on mapping class groups". *Ann. Inst. Fourier (Grenoble)* 66.3, pp. 871–898. MR: 3494163 (cit. on p. 1123).

Mladen Bestvina and Kenneth Bromberg (n.d.). "On the asymptotic dimension of a curve complex". arXiv: 1508.04832 (cit. on p. 1120).

Mladen Bestvina, Kenneth Bromberg, and Koji Fujiwara (n.d.). "Bounded cohomology via quasi-trees". arXiv: 1306.1542 (cit. on p. 1124).

Mladen Bestvina, Kenneth Bromberg, Koji Fujiwara, and Alessandro Sisto (n.d.). "Acylindrical actions on projection complexes". in preparation (cit. on p. 1127).

Mladen Bestvina and Mark Feighn (2014a). "Hyperbolicity of the complex of free factors". *Advances in Mathematics* 256, pp. 104–155 (cit. on p. 1128).

– (2014b). "Subfactor projections". *Journal of Topology*. 34 pages (cit. on p. 1128).

Mladen Bestvina and Koji Fujiwara (2002). "Bounded cohomology of subgroups of mapping class groups". *Geom. Topol.* 6, 69–89 (electronic). MR: 1914565 (cit. on pp. 1124, 1125, 1128).

– (2007). "Quasi-homomorphisms on mapping class groups". *Glas. Mat. Ser. III* 42(62).1, pp. 213–236. MR: 2332668 (cit. on p. 1117).

– (2009). "A characterization of higher rank symmetric spaces via bounded cohomology". *Geom. Funct. Anal.* 19.1, pp. 11–40. MR: 2507218 (cit. on pp. 1124, 1125).

Mladen Bestvina, Vincent Guirardel, and Camille Horbez (n.d.). "Boundary amenability of $Out(F_n)$". arXiv: 1705.07017 (cit. on p. 1129).

Brian H. Bowditch (2008). "Tight geodesics in the curve complex". *Invent. Math.* 171.2, pp. 281–300. MR: 2367021 (cit. on p. 1127).

Emmanuel Breuillard and Koji Fujiwara (n.d.). "On the joint spectral radius for isometries of non-positively curved spaces and uniform growth". arXiv: 1804.00748 (cit. on pp. 1121, 1122).

Martin R. Bridson (2010). "Semisimple actions of mapping class groups on CAT(0) spaces". In: *Geometry of Riemann surfaces*. Vol. 368. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press, pp. 1–14. MR: 2665003 (cit. on p. 1121).

Martin R. Bridson and Karen Vogtmann (1995). "On the geometry of the automorphism group of a free group". *Bull. London Math. Soc.* 27.6, pp. 544–552. MR: 1348708 (cit. on p. 1128).

Robert Brooks (1981). "Some remarks on bounded cohomology". In: *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978)*. Vol. 97. Ann. of Math. Stud. Princeton, N.J.: Princeton Univ. Press, pp. 53–63. MR: 624804 (cit. on p. 1122).

Marc Burger and Shahar Mozes (2000). "Lattices in product of trees". *Inst. Hautes Études Sci. Publ. Math.* 92, 151–194 (2001). MR: 1839489 (cit. on p. 1127).

Danny Calegari (2009). *scl*. Vol. 20. MSJ Memoirs. Mathematical Society of Japan, Tokyo, pp. xii+209. MR: 2527432 (cit. on p. 1122).

Danny Calegari and Koji Fujiwara (2010). "Stable commutator length in word-hyperbolic groups". *Groups Geom. Dyn.* 4.1, pp. 59–90. MR: 2566301 (cit. on p. 1123).

Serge Cantat and Stéphane Lamy (2013). "Normal subgroups in the Cremona group". *Acta Math.* 210.1. With an appendix by Yves de Cornulier, pp. 31–94. MR: 3037611 (cit. on p. 1126).

Pierre-Emmanuel Caprace and Koji Fujiwara (2010). "Rank-one isometries of buildings and quasi-morphisms of Kac-Moody groups". *Geom. Funct. Anal.* 19.5, pp. 1296–1319. MR: 2585575 (cit. on p. 1125).

Pierre-Emmanuel Caprace and Michah Sageev (2011). "Rank rigidity for CAT(0) cube complexes". *Geom. Funct. Anal.* 21.4, pp. 851–891. MR: 2827012 (cit. on pp. 1125, 1127).

Ruth Charney and Harold Sultan (2015). "Contracting boundaries of CAT(0) spaces". *J. Topol.* 8.1, pp. 93–117. MR: 3339446 (cit. on p. 1125).

Marc Culler and Karen Vogtmann (1986). "Moduli of graphs and automorphisms of free groups". *Invent. Math.* 84.1, pp. 91–119. MR: 830040 (cit. on p. 1126).

F. Dahmani, V. Guirardel, and D. Osin (2017). "Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces". *Mem. Amer. Math. Soc.* 245.1156, pp. v+152. MR: 3589159 (cit. on p. 1126).

Thomas Delzant (2016). "Kähler groups, $\mathbb{R}$-trees, and holomorphic families of Riemann surfaces". *Geom. Funct. Anal.* 26.1, pp. 160–187. MR: 3494488 (cit. on p. 1124).

Cornelia Druţu and Mark Sapir (2005). "Tree-graded spaces and asymptotic cones of groups". *Topology* 44.5. With an appendix by Denis Osin and Mark Sapir, pp. 959–1058. MR: 2153979 (cit. on p. 1119).

H. Endo and D. Kotschick (2001). "Bounded cohomology and non-uniform perfection of mapping class groups". *Invent. Math.* 144.1, pp. 169–175. MR: 1821147 (cit. on p. 1123).

– (2007). "Failure of separation by quasi-homomorphisms in mapping class groups". *Proc. Amer. Math. Soc.* 135.9, pp. 2747–2750. MR: 2317948 (cit. on p. 1123).

David B. A. Epstein and Koji Fujiwara (1997). "The second bounded cohomology of word-hyperbolic groups". *Topology* 36.6, pp. 1275–1289. MR: 1452851 (cit. on pp. 1123, 1128).

Benson Farb, Alexander Lubotzky, and Yair N. Minsky (2001). "Rank-1 phenomena for mapping class groups". *Duke Math. J.* 106.3, pp. 581–597. MR: 1813237 (cit. on p. 1120).

Benson Farb and Dan Margalit (2012). *A primer on mapping class groups*. Vol. 49. Princeton Mathematical Series. Princeton, NJ: Princeton University Press, pp. xiv+472. MR: 2850125 (cit. on p. 1116).

Koji Fujiwara (1998). "The second bounded cohomology of a group acting on a Gromov-hyperbolic space". *Proc. London Math. Soc. (3)* 76.1, pp. 70–94. MR: 1476898 (cit. on p. 1128).

– (2000). "The second bounded cohomology of an amalgamated free product of groups". *Trans. Amer. Math. Soc.* 352.3, pp. 1113–1129. MR: 1491864 (cit. on p. 1128).

M. Gromov (1987). "Hyperbolic groups". In: *Essays in group theory*. Vol. 8. Math. Sci. Res. Inst. Publ. New York: Springer, pp. 75–263. MR: 919829 (cit. on p. 1105).

– (1993). "Asymptotic invariants of infinite groups". In: *Geometric group theory, Vol. 2 (Sussex, 1991)*. Vol. 182. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press, pp. 1–295. MR: 1253544 (cit. on p. 1109).

Thomas Haettel (n.d.). "Hyperbolic rigidity of higher rank lattices". arXiv: 1607.02004 (cit. on p. 1128).

Ursula Hamenstädt (2009). "Geometry of the mapping class groups. I. Boundary amenability". *Invent. Math.* 175.3, pp. 545–609. MR: 2471596 (cit. on p. 1110).

Michael Handel and Lee Mosher (n.d.). "Hyperbolic actions and 2nd bounded cohomology of subgroups of $Out(F_n)$. Part II: Finite lamination subgroups". arXiv: 1702.08050 (cit. on p. 1124).

– (2013). "The free splitting complex of a free group, I: hyperbolicity". *Geom. Topol.* 17.3, pp. 1581–1672. MR: 3073931 (cit. on p. 1128).

Pierre de la Harpe and Alain Valette (1989). "La propriété (T) de Kazhdan pour les groupes localement compacts (avec un appendice de Marc Burger)". *Astérisque* 175. With an appendix by M. Burger, p. 158. MR: 1023471 (cit. on p. 1125).

Sebastian Hensel, Piotr Przytycki, and Richard C. H. Webb (2015). "1-slim triangles and uniform hyperbolicity for arc graphs and curve graphs". *J. Eur. Math. Soc. (JEMS)* 17.4, pp. 755–762. MR: 3336835 (cit. on p. 1117).

Yoshikata Kida (2008). "The mapping class group from the viewpoint of measure equivalence theory". *Mem. Amer. Math. Soc.* 196.916, pp. viii+190. MR: 2458794 (cit. on p. 1110).

Mustafa Korkmaz (2004). "Stable commutator length of a Dehn twist". *Michigan Math. J.* 52.1, pp. 23–31. MR: 2043394 (cit. on p. 1123).

Johanna Mangahas (2010). "Uniform uniform exponential growth of subgroups of the mapping class group". *Geom. Funct. Anal.* 19.5, pp. 1468–1480. MR: 2585580 (cit. on pp. 1118, 1121).

– (2013). "A recipe for short-word pseudo-Anosovs". *Amer. J. Math.* 135.4, pp. 1087–1116. MR: 3086070 (cit. on p. 1118).

J. F. Manning (2006). "Quasi-actions on trees and property (QFA)". *J. London Math. Soc. (2)* 73.1. With an appendix by N. Monod and B. Rémy, pp. 84–108. MR: 2197372 (cit. on pp. 1120, 1128).

Jason Fox Manning (2005). "Geometry of pseudocharacters". *Geom. Topol.* 9, 1147–1185 (electronic). MR: 2174263 (cit. on pp. 1109, 1128).

H. A. Masur and Yair N. Minsky (2000). "Geometry of the complex of curves. II. Hierarchical structure". *Geom. Funct. Anal.* 10.4, pp. 902–974. MR: 1791145 (cit. on pp. 1116–1119).

Howard A. Masur and Yair N. Minsky (1999). "Geometry of the complex of curves. I. Hyperbolicity". *Invent. Math.* 138.1, pp. 103–149. MR: 1714338 (cit. on p. 1116).

Yair N. Minsky (1996). "Quasi-projections in Teichmüller space". *J. Reine Angew. Math.* 473, pp. 121–136. MR: 1390685 (cit. on pp. 1116, 1126).

– (2006). "Curve complexes, surfaces and 3-manifolds". In: *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, pp. 1001–1033. MR: 2275633 (cit. on p. 1116).

Nicolas Monod (2006). "An invitation to bounded cohomology". In: *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, pp. 1183–1211. MR: 2275641 (cit. on p. 1128).

Lee Mosher, Michah Sageev, and Kevin Whyte (2003). "Quasi-actions on trees. I. Bounded valence". *Ann. of Math. (2)* 158.1, pp. 115–164. MR: 1998479 (cit. on p. 1121).

D. Osin (2016). "Acylindrically hyperbolic groups". *Trans. Amer. Math. Soc.* 368.2, pp. 851–888. MR: 3430352 (cit. on p. 1127).

John Roe (2003). *Lectures on coarse geometry*. Vol. 31. University Lecture Series. American Mathematical Society, Providence, RI, pp. viii+175. MR: 2007488 (cit. on pp. 1110, 1120).

Z. Sela (1997). "Acylindrical accessibility for groups". *Invent. Math.* 129.3, pp. 527–565. MR: 1465334 (cit. on p. 1126).

Jean-Pierre Serre (1980). *Trees*. Translated from the French by John Stillwell. Springer-Verlag, Berlin-New York, pp. ix+142. MR: 607504 (cit. on p. 1105).

Karen Vogtmann (2006). "The cohomology of automorphism groups of free groups". In: *International Congress of Mathematicians. Vol. II*. Eur. Math. Soc., Zürich, pp. 1101–1117. MR: 2275637 (cit. on p. 1126).

Richard C. H. Webb (2015). "Combinatorics of tight geodesics and stable lengths". *Trans. Amer. Math. Soc.* 367.10, pp. 7323–7342. MR: 3378831 (cit. on p. 1120).

Guoliang Yu (1998). "The Novikov conjecture for groups with finite asymptotic dimension". *Ann. of Math. (2)* 147.2, pp. 325–355. MR: 1626745 (cit. on p. 1110).

Received 2017-11-21.

Koji Fujiwara (藤原　耕二)
Department of Mathematics
Kyoto University
Japan
kfujiwara@math.kyoto-u.ac.jp

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1133–1170)

GEOMETRIC STRUCTURES AND REPRESENTATIONS OF DISCRETE GROUPS

Fanny Kassel

Abstract

We describe recent links between two topics: geometric structures on manifolds in the sense of Ehresmann and Thurston, and dynamics "at infinity" for representations of discrete groups into Lie groups.

1 Introduction

The goal of this survey is to report on recent results relating geometric structures on manifolds to dynamical aspects of representations of discrete groups into Lie groups, thus linking geometric topology to group theory and dynamics.

1.1 Geometric structures. The first topic of this survey is geometric structures on manifolds. Here is a concrete example as illustration (see Figure 1).

Example 1.1. Consider a two-dimensional torus T.

(1) We can view T as the quotient of the Euclidean plane $X = \mathbb{R}^2$ by $\Gamma = \mathbb{Z}^2$, which is a discrete subgroup of the isometry group $G = \mathrm{O}(2) \ltimes \mathbb{R}^2$ of X (acting by linear isometries and translations). Viewing T this way provides it with a Riemannian metric and a notion of parallel lines, length, angles, etc. We say T is endowed with a *Euclidean* (or *flat*) *structure*, or a (G, X)-structure with $(G, X) = (\mathrm{O}(2) \ltimes \mathbb{R}^2, \mathbb{R}^2)$.

(2) Here is a slightly more involved way to view T: we can see it as the quotient of the affine plane $X = \mathbb{R}^2$ by the group Γ generated by the translation of vector $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and the affine transformation with linear part $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and translational part $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$. This group Γ is now a discrete subgroup of the affine group $G = \mathrm{GL}(2, \mathbb{R}) \ltimes \mathbb{R}^2$. Viewing T this way still provides it with a notion of parallel lines and even of geodesic, but no longer with a notion

Partially supported by the European Research Council under Starting Grant 715982 (DiGGeS).
MSC2010: primary 57N16; secondary 20H10, 22E40, 37D40, 57M50, 57S30.

of length or angle or speed of geodesic. We say T is endowed with an *affine structure*, or a (G, X)-structure with $(G, X) = (\mathrm{GL}(2, \mathbb{R}) \ltimes \mathbb{R}^2, \mathbb{R}^2)$.

(3) There are many ways to endow T with an affine structure. Here is a different one: we can view T as the quotient of the open subset $\mathcal{U} = \mathbb{R}^2 \smallsetminus \{0\}$ of $X = \mathbb{R}^2$ by the discrete subgroup Γ of $G = \mathrm{GL}(2, \mathbb{R}) \ltimes \mathbb{R}^2$ generated by the homothety $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$. This still makes T "locally look like" $X = \mathbb{R}^2$, but now the image in T of an affine geodesic of X pointing towards the origin is *incomplete* (it circles around in T with shorter and shorter period and disappears in a finite amount of time).

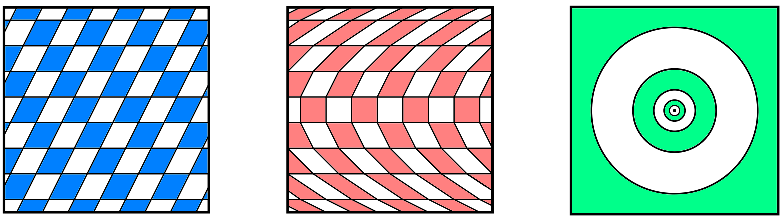


Figure 1: Tilings of $X = \mathbb{R}^2$ showing the three Γ-actions in Example 1.1

As in Example 1.1, a key idea underlying a large part of modern geometry is the existence of *model geometries* which various manifolds may locally carry. By definition, a model geometry is a pair (G, X) where X is a manifold (*model space*) and G a Lie group acting transitively on X (*group of symmetries*). In Example 1.1 we encountered $(G, X) = (\mathrm{O}(n) \ltimes \mathbb{R}^n, \mathbb{R}^n)$ and $(G, X) = (\mathrm{GL}(n, \mathbb{R}) \ltimes \mathbb{R}^n, \mathbb{R}^n)$, corresponding respectively to *Euclidean geometry* and *affine geometry*. Another important example is $X = \mathbb{H}^n$ (the n-dimensional real hyperbolic space) and $G = \mathrm{PO}(n, 1) = \mathrm{O}(n, 1)/\{\pm I\}$ (its group of isometries), corresponding to *hyperbolic geometry*. (For $n = 2$ we can see X as the upper half-plane and G, up to index two, as $\mathrm{PSL}(2, \mathbb{R})$ acting by homographies.) We refer to Table 1 for more examples.

The idea that a manifold M locally carries the geometry (G, X) is formalized by the notion of a (G, X)-*structure* on M: by definition, this is a maximal atlas of coordinate charts on M with values in X such that the transition maps are given by elements of G (see Figure 2). Note that this is quite similar to a manifold structure on M, but we now require the charts to take values in X rather than $\mathbb{R}^n$, and the transition maps to be given by elements of G rather than diffeomorphisms of $\mathbb{R}^n$. Although general (G, X)-structures may display pathological behavior (see Goldman [2018b]), in this survey we will restrict to the two "simple" types of (G, X)-structures appearing in Example 1.1, to which we shall give names to facilitate the discussion:

- **Type C ("complete")**: (G, X)-structures that identify M with a quotient of X by a discrete subgroup Γ of G acting properly discontinuously;

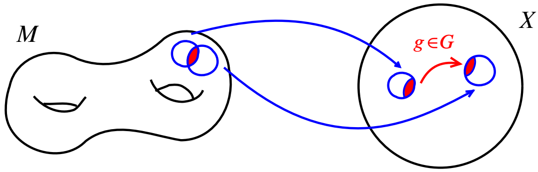


Figure 2: Charts defining a (G, X)-structure on M

- **Type U ("incomplete but still uniformizable")**: (G, X)-structures that identify M with a quotient of some proper open subset $\mathcal{U}$ of X by a discrete subgroup Γ of G acting properly discontinuously.

Setting $\mathcal{V} = X$ or $\mathcal{U}$ as appropriate, we then have coverings $\widetilde{M} \simeq \widetilde{\mathcal{V}} \to \mathcal{V} \to \Gamma\backslash\mathcal{V} \simeq M$ (where $\widetilde{\ }$ denotes universal covers). The charts on M are obtained by taking preimages in $\mathcal{V} \subset X$ of open subsets of M. Moreover, the basic theory of covering groups gives a natural group homomorphism $\mathrm{hol} : \pi_1(M) \to G$ with image Γ and kernel $\pi_1(\mathcal{V})$, called the *holonomy*.

In this survey, we use the phrase *geometric structures* for (G, X)-structures. We shall not detail the rich historical aspects of geometric structures here; instead, we refer to the excellent surveys of Goldman [2010, 2018a,b]. Let us just mention that the notion of model geometry has its origins in ideas of Lie and Klein, formulated in Klein's 1872 Erlangen program. Influenced by these ideas and those of Poincaré, Cartan and others, Ehresmann [1937] initiated a general study of geometric structures in the 1930s. Later, geometric structures were greatly promoted by the revolutionary work of Thurston [1980].

1.2 Classifying geometric structures. The fundamental problem in the theory of geometric structures is their classification, namely:

Problem A. Given a manifold M,
(1) Describe which model geometries (G, X) the manifold M may locally carry;
(2) For a fixed model (G, X), describe all possible (G, X)-structures on M.

We refer to Goldman [2010] for a detailed survey of Problem A with a focus on dimensions two and three, and to Kobayashi and Yoshino [2005] for a special case.

Problem A.(1) asks how the global topology of M determines the geometries that it may locally carry. This has been the object of deep results, among which:

- the classical *uniformization theorem*: a closed Riemann surface may carry a Euclidean, a spherical, or a hyperbolic structure, depending on its genus;

- Thurston's *hyperbolization theorem*: a large class of 3-dimensional manifolds, defined in purely topological terms, may carry a hyperbolic structure;

- more generally, Thurston's *geometrization program* (now Perelman's theorem): any closed orientable 3-dimensional manifold may be decomposed into pieces, each admitting one of eight model geometries (see Bonahon [2002]).

Problem A.(2) asks to describe the *deformation space* of (G, X)-structures on M. In the simple setting of Example 1.1, this space is already quite rich (see Baues [2014]). For hyperbolic structures on a closed Riemann surface of genus ≥ 2 (Example 2.1), Problem A.(2) gives rise to the fundamental and wide-ranging *Teichmüller theory*.

1.3 Representations of discrete groups. The second topic of this survey is representations (i.e. group homomorphisms) of discrete groups (i.e. countable groups) to Lie groups G, and their dynamics "at infinity". We again start with an example.

Example 1.2. Let $\Gamma = \pi_1(S)$ where S is a closed orientable Riemann surface of genus ≥ 2. By the uniformization theorem, S carries a complete ("type C") hyperbolic structure, which yields a holonomy representation $\Gamma \to \mathrm{PSL}(2, \mathbb{R})$ as in Section 1.1. Embedding $\mathrm{PSL}(2, \mathbb{R})$ into $G = \mathrm{PSL}(2, \mathbb{C})$, we obtain a representation $\rho : \Gamma \to G$, called *Fuchsian*, and an associated action of Γ on the hyperbolic space $X = \mathbb{H}^3$ and on its boundary at infinity $\partial_\infty \mathbb{H}^3 = \widehat{\mathbb{C}}$ (the Riemann sphere). The *limit set* of $\rho(\Gamma)$ in $\widehat{\mathbb{C}}$ is the set of accumulation points of $\rho(\Gamma)$-orbits of X; it is a circle in the sphere $\widehat{\mathbb{C}}$. Deforming ρ slightly yields a new representation $\rho' : \Gamma \to G$, called *quasi-Fuchsian*, which is still faithful, with discrete image, and whose limit set in $\widehat{\mathbb{C}}$ is still a topological circle (now "wiggly", see Figure 3). The action of $\rho'(\Gamma)$ is chaotic on the limit set (e.g. all orbits are dense) and properly discontinuous on its complement.

Example 1.2 plays a central role in the theory of Kleinian groups and in Thurston's geometrization program; it was extensively studied by Ahlfors, Beardon, Bers, Marden, Maskit, Minsky, Sullivan, Thurston, and many others.

In this survey we report on various generalizations of Example 1.2, for representations of discrete groups Γ to semisimple Lie groups G which are faithful (or with finite kernel) and whose images are discrete subgroups of G. While in Example 1.2 the group $G = \mathrm{PSL}(2, \mathbb{C})$ has real rank one (meaning that its Riemannian symmetric space $\mathbb{H}^3$ has no flat regions beyond geodesics), we also wish to consider the case that G has *higher real rank*, e.g. $G = \mathrm{PGL}(d, \mathbb{R})$ with $d \geq 3$. In general, semisimple groups G tend to have very different behavior depending on whether their real rank is one or higher; for instance, the *lattices* of G (i.e. the discrete subgroups of finite covolume for the Haar measure) may display some forms of flexibility in real rank one, but exhibit strong rigidity phenomena in higher real rank. Beyond lattices, the landscape of discrete subgroups of G is somewhat understood in real rank one (at least several important classes of discrete subgroups

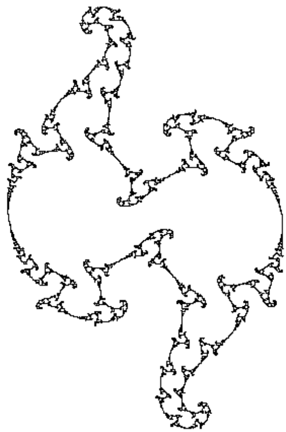


Figure 3: The limit set of a quasi-Fuchsian group in $\partial_\infty \mathbb{H}^3 \simeq \mathbb{C} \cup \{\infty\}$

have been identified for their good geometric, topological, and dynamical properties, see Section 3.1), but it remains very mysterious in higher real rank. We shall explain some recent attempts at understanding it better.

One interesting aspect is that, even when G has higher real rank, discrete subgroups of G of infinite covolume may be nonrigid and in fact admit large deformation spaces. In particular, as part of *higher Teichmüller theory*, there has recently been an active and successful effort to find large deformation spaces of faithful and discrete representations of surface groups $\pi_1(S)$ into higher-rank semisimple G which share some of the rich features of the Teichmüller space of S (see Sections 4.3 and 5, Burger, Iozzi, and Wienhard [2014] and Wienhard [2018]). Such features also include dynamics "at infinity" as in Example 1.2, which are encompassed by a notion of *Anosov representation* (Labourie [2006]), see Section 4.

1.4 Flag varieties and boundary maps. Let us be a bit more precise. Given a representation $\rho : \Gamma \to G$, by dynamics "at infinity" we mean the dynamics of the action of Γ via ρ on some *flag varieties* G/P (where P is a parabolic subgroup), seen as "boundaries" of G or of its Riemannian symmetric space G/K. In Example 1.2 we considered a rank-one situation where $G = \mathrm{PSL}(2, \mathbb{C})$ and $G/P = \partial_\infty \mathbb{H}^3 = \widehat{\mathbb{C}}$. A typical higher-rank situation that we have in mind is $G = \mathrm{PGL}(d, \mathbb{R})$ with $d \geq 3$ and $G/P = \mathrm{Gr}_i(\mathbb{R}^d)$ (the Grassmannian of i-planes in $\mathbb{R}^d$) for some $1 \leq i \leq d-1$.

In the work of Mostow, Margulis, Furstenberg, and others, rigidity results have often relied on the construction of Γ-equivariant measurable maps from or to G/P. More recently, in the context of higher Teichmüller theory (see Burger, Iozzi, and Wienhard [2010b], Fock and Goncharov [2006], Labourie [2006]), it has proved important to study continuous equivariant *boundary maps* which embed the boundary $\partial_\infty\Gamma$ of a Gromov hyperbolic group Γ (i.e. the visual boundary of the Cayley graph of Γ) into G/P. Such boundary maps $\xi : \partial_\infty\Gamma \to G/P$ define a closed invariant subset $\xi(\partial_\infty\Gamma)$ of G/P, the *limit set*, on which the dynamics of the action by Γ accurately reflect the intrinsic chaotic dynamics of Γ on $\partial_\infty\Gamma$. These boundary maps may be used to transfer the Anosov property of the intrinsic geodesic flow of Γ into some uniform contraction/expansion properties for a flow on a natural flat bundle associated to ρ and G/P (see Section 4). They may also define some open subsets $\mathcal{U}$ of G/P on which the action of Γ is properly discontinuous, by removing an "extended limit set" $\mathcal{L}_{\rho(\Gamma)} \supset \xi(\partial_\infty\Gamma)$ (see Sections 3, 5 and 6); this generalizes the domains of discontinuity in the Riemann sphere of Example 1.2.

For finitely generated groups Γ that are not Gromov hyperbolic, one can still define a boundary $\partial_\infty\Gamma$ in several natural settings, e.g. as the visual boundary of some geodesic metric space on which Γ acts geometrically, and the approach considered in this survey can then be summarized by the following general problem.

Problem B. Given a discrete group Γ with a boundary $\partial_\infty\Gamma$, and a Lie group G with a boundary G/P, identify large (e.g. open in $\mathrm{Hom}(\Gamma, G)$) classes of faithful and discrete representations $\rho : \Gamma \to G$ for which there exist continuous ρ-equivariant boundary maps $\xi : \partial_\infty\Gamma \to G/P$. Describe the dynamics of Γ on G/P via ρ.

1.5 Goal of the paper. We survey recent results on (G, X)-structures (Problem A) and on representations of discrete groups (Problem B), making links between the two topics. In one direction, we observe that various types of (G, X)-structures have holonomy representations that are interesting for Problem B. In the other direction, starting with representations that are interesting for Problem B (Anosov representations), we survey recent constructions of associated (G, X)-structures. These results tend to indicate some deep interactions between the geometry of (G, X)-manifolds and the dynamics of their holonomy representations, which largely remain to be explored. We hope that they will continue to stimulate the development of rich theories in the future.

Organization of the paper. In Section 2 we briefly review the notion of a holonomy representation. In Section 3 we describe three important families of (G, X)-structures for which boundary maps into flag varieties naturally appear. In Section 4 we define Anosov representations and give examples and characterizations. In Section 5 we summarize recent constructions of geometric structures associated to Anosov representations. In Section 6 we discuss a situation in which the links between geometric structures and Anosov representations are particularly tight, in the context of convex projective geometry. In Section 7 we examine an instance of (G, X)-structures for a nonreductive Lie group G, corresponding to affine manifolds and giving rise to affine Anosov representations. We conclude with a few remarks.

Acknowledgments. I would like to heartily thank all the mathematicians who helped, encouraged, and inspired me over the past years; the list is too long to include here. I am very grateful to all my coauthors, in particular those involved in the work discussed below: Jeffrey Danciger (Sections 5 to 7), François Guéritaud (Sections 4 to 7), Olivier Guichard (Sections 4 and 5), Rafael Potrie (Section 4), and Anna Wienhard (Sections 4 and 5). I warmly thank J.-P. Burelle, J. Danciger, O. Guichard, and S. Maloni for reading earlier versions of this text and making many valuable comments and suggestions, and R. Canary and W. Goldman for kindly answering my questions.

2 Holonomy representations

Let G be a real Lie group acting transitively, faithfully, analytically on a manifold X, as in Table 1. In Section 1.1 we defined holonomy representations for certain types of (G, X)-structures. We now give a short review of the notion in general.

Type of geometry	X	G	H
Real projective	$\mathbb{P}^n(\mathbb{R})$	$\mathrm{PGL}(n+1, \mathbb{R})$	stab. of line of $\mathbb{R}^{n+1}$
Affine	$\mathbb{R}^n$	$\mathrm{Aff}(\mathbb{R}^n) = \mathrm{GL}(n, \mathbb{R}) \ltimes \mathbb{R}^n$	$\mathrm{GL}(n, \mathbb{R})$
Euclidean	$\mathbb{R}^n$	$\mathrm{Isom}(\mathbb{R}^n) = \mathrm{O}(n) \ltimes \mathbb{R}^n$	$\mathrm{O}(n)$
Real hyperbolic	$\mathbb{H}^n$	$\mathrm{Isom}(\mathbb{H}^n) = \mathrm{PO}(n, 1)$	$\mathrm{O}(n)$
Spherical	$\mathbb{S}^n$	$\mathrm{Isom}(\mathbb{S}^n) = \mathrm{O}(n+1)$	$\mathrm{O}(n)$
Complex projective	$\mathbb{P}^n(\mathbb{C})$	$\mathrm{PGL}(n+1, \mathbb{C})$	stab. of line of $\mathbb{C}^{n+1}$

Table 1: Some examples of model geometries (G, X), where $X \simeq G/H$

Let M be a (G, X)-manifold, i.e. a manifold endowed with a (G, X)-structure. Fix a basepoint $m \in M$ and a chart $\varphi : \mathcal{U} \to X$ with $m \in \mathcal{U}$. We can lift any loop on M starting at m to a path on X starting at $\varphi(m)$, using successive charts of M which coincide on their intersections; the last chart in this analytic continuation process coincides, on an open set, with $g \cdot \varphi$ for some unique $g \in G$; we set $\mathrm{hol}(\gamma) := g$ where $\gamma \in \pi_1(M, m)$ is the homotopy class of the loop (see Figure 4). This defines a representation $\mathrm{hol} : \pi_1(M) \to G$ called the *holonomy* (see Goldman [2010, 2018b] for details); it is unique modulo conjugation by G. This coincides with the notion from Section 1.1; in particular, if $M \simeq \Gamma\backslash\mathcal{V}$ with $\mathcal{V}$ open in X and Γ discrete in G, and if $\mathcal{V}$ is simply connected, then $\mathrm{hol} : \pi_1(M) \to \Gamma$ is just the natural identification of $\pi_1(M)$ with Γ.

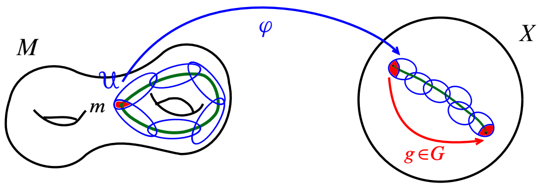


Figure 4: Construction of a holonomy representation

The *deformation space* $\mathrm{Def}_{(G,X)}(M)$ of (G,X)-structures on M is the quotient of the set of (G,X)-structures by the group of diffeomorphisms of M homotopic to the identity, acting by precomposition of the charts. It has a natural topology, see Baues [2014, § 3]. The holonomy defines a map from $\mathrm{Def}_{(G,X)}(M)$ to the space $\mathrm{Hom}(\Gamma, G)/G$ of representations of Γ to G modulo conjugation by G. This map may be injective in some cases, as in Example 2.1 below, but in general it is not. However, when M is closed, the so-called *Ehresmann–Thurston principle* (see Thurston [1980]) states that the map is continuous, open, with discrete fibers; in particular, the set of holonomy representations of (G,X)-structures on M is then stable under small deformations.

Example 2.1. Let $(G,X) = (\mathrm{PO}(2,1), \mathbb{H}^2)$ where $\mathrm{PO}(2,1) \simeq \mathrm{PGL}(2,\mathbb{R})$ is the isometry group of the real hyperbolic plane $\mathbb{H}^2$. Let $M = S$ be a closed orientable connected surface of genus $g \geq 2$. All (G,X)-structures on S are complete. The deformation space $\mathrm{Def}_{(G,X)}(S)$ is the Teichmüller space $\mathrm{Teich}(S)$ of S. The holonomy defines a homeomorphism between $\mathrm{Teich}(S) \simeq \mathbb{R}^{6g-6}$ and the space of Fuchsian (i.e. faithful and discrete) representations from $\pi_1(S)$ to G modulo conjugation by G.

3 Examples of $(\mathbf{G},\mathbf{X})$-structures and their holonomy representations

In this section we introduce three important families of (G,X)-structures, which have been much studied in the past few decades. We observe some structural stability for their holonomy representations, and the existence of continuous equivariant boundary maps together with expansion/contraction properties "at infinity". These phenomena will be captured by the notion of an Anosov representation in Section 4.

3.1 Convex cocompact locally symmetric structures in rank one. Let G be a real semisimple Lie group of real rank one with Riemannian symmetric space $X = G/K$ (i.e. K is a maximal compact subgroup of G), e.g. $(G,X) = (\mathrm{PO}(n,1), \mathbb{H}^n)$ for $n \geq 2$. Convex cocompact groups are an important class of discrete subgroups Γ of G which generalize the uniform lattices. They are special cases of geometrically finite groups, for which no cusps appear; see Bowditch [1993, 1998] for a general theory.

By definition, a discrete subgroup Γ of G is *convex cocompact* if it preserves and acts with compact quotient on some nonempty convex subset $\mathcal{C}$ of $X = G/K$; equivalently, the complete (G, X)-manifold (or orbifold) $\Gamma\backslash X$ has a compact convex subset (namely $\Gamma\backslash\mathcal{C}$) containing all the topology. Such a group Γ is always finitely generated. A representation $\rho : \Gamma \to G$ is called *convex cocompact* if its kernel is finite and its image is a convex cocompact subgroup of G.

For instance, in Example 1.2 the quasi-Fuchsian representations are exactly the convex cocompact representations from $\pi_1(S)$ to $G = \mathrm{PSL}(2, \mathbb{C})$; modulo conjugation, they are parametrized by $\mathrm{Teich}(S) \times \mathrm{Teich}(S)$ (Bers [1960]). Another classical example of convex cocompact groups in rank-one G is Schottky groups, namely free groups defined by the so-called *ping pong* dynamics of their generators in $\partial_\infty X$.

Here $\partial_\infty X$ denotes the visual boundary of X, yielding the standard compactification $\overline{X} = X \sqcup \partial_\infty X$ of X; for $X = \mathbb{H}^n$ we can see $\overline{X}$ in projective space as in Example 3.2.(1) below. The G-action on X extends continuously to $\overline{X}$, and $\partial_\infty X$ identifies with G/P where P is a minimal parabolic subgroup of G.

For a convex cocompact representation $\rho : \Gamma \to G$, the existence of a cocompact invariant convex set $\mathcal{C}$ implies (by the Švarc–Milnor lemma or "fundamental observation of geometric group theory") that ρ is a *quasi-isometric embedding*. This means that the points of any $\rho(\Gamma)$-orbit in $X = G/K$ go to infinity at linear speed for the word length function $|\cdot| : \Gamma \to \mathbb{N}$: for any $x_0 \in X$ there exist $C, C' > 0$ such that $d_{G/K}(x_0, \rho(\gamma) \cdot x_0) \geq C\,|\gamma| - C'$ for all $\gamma \in \Gamma$. (This property does not depend on the choice of finite generating subset of Γ defining $|\cdot|$.) A consequence "at infinity" is that any ρ-orbital map $\Gamma \to X$ extends to a ρ-equivariant embedding $\xi : \partial_\infty\Gamma \to \partial_\infty X \simeq G/P$, where $\partial_\infty\Gamma$ is the boundary of the Gromov hyperbolic group Γ. The image of ξ is the *limit set* $\Lambda_{\rho(\Gamma)}$ of $\rho(\Gamma)$ in $\partial_\infty X$. The dynamics on $\partial_\infty X \simeq G/P$ are decomposed as in Example 1.2: the action of $\rho(\Gamma)$ is "chaotic" on $\Lambda_{\rho(\Gamma)}$ (e.g. all orbits are dense if Γ is nonelementary), and properly discontinuous, with compact quotient, on the complement $\Omega_{\rho(\Gamma)} = \partial_\infty X \smallsetminus \Lambda_{\rho(\Gamma)}$.

Further dynamical properties were described by Sullivan [1979, 1985]: for instance, the action of $\rho(\Gamma)$ on $\partial_\infty X \simeq G/P$ is *expanding* at each point $z \in \Lambda_{\rho(\Gamma)}$, i.e. there exist $\gamma \in \Gamma$ and $C > 1$ such that $\rho(\gamma)$ multiplies all distances by $\geq C$ on a neighborhood of z in $\partial_\infty X$ (for some fixed auxiliary metric on $\partial_\infty X$). This implies that the group $\rho(\Gamma)$ is *structurally stable*, i.e. there is a neighborhood of the natural inclusion in $\mathrm{Hom}(\rho(\Gamma), G)$ consisting entirely of faithful representations. In fact, ρ admits a neighborhood consisting entirely of convex cocompact representations, by a variant of the Ehresmann–Thurston principle. For $G = \mathrm{SL}(2, \mathbb{C})$, a structurally stable subgroup of G is either locally rigid or convex cocompact (Sullivan [1985]).

3.2 Convex projective structures: divisible convex sets. Let G be the projective linear group $\mathrm{PGL}(d,\mathbb{R})$ and X the projective space $\mathbb{P}(\mathbb{R}^d)$, for $d \geq 2$. Recall that a subset of $X = \mathbb{P}(\mathbb{R}^d)$ is said to be *convex* if it is contained and convex in some affine chart, *properly convex* if its closure is convex, and *strictly convex* if it is properly convex and its boundary in X does not contain any nontrivial segment.

Remark 3.1. Any properly convex open subset Ω of $X = \mathbb{P}(\mathbb{R}^d)$ admits a well-behaved (complete, proper, Finsler) metric d_Ω, the *Hilbert metric*, which is invariant under the subgroup of $G = \mathrm{PGL}(d,\mathbb{R})$ preserving Ω (see e.g. Benoist [2008]). In particular, any discrete subgroup of G preserving Ω acts properly discontinuously on Ω.

By definition, a *convex projective structure* on a manifold M is a (G,X)-structure obtained by identifying M with $\Gamma\backslash\Omega$ for some properly convex open subset Ω of X and some discrete subgroup Γ of G. When M is closed, i.e. when Γ acts with compact quotient, we say that Γ *divides* Ω. Such *divisible convex sets* Ω are the objects of a rich theory, see Benoist [ibid.]. The following classical examples are called *symmetric*.

Examples 3.2. (1) For $d = n+1 \geq 3$, let $\langle\cdot,\cdot\rangle_{n,1}$ be a symmetric bilinear form of signature $(n,1)$ on $\mathbb{R}^d$, and let $\Omega = \{[v] \in \mathbb{P}(\mathbb{R}^d) \mid \langle v,v\rangle_{n,1} < 0\}$ be the projective model of the real hyperbolic space $\mathbb{H}^n$. It is a strictly convex open subset of $X = \mathbb{P}(\mathbb{R}^d)$ (an ellipsoid), and any uniform lattice Γ of $\mathrm{PO}(n,1) \subset G = \mathrm{PGL}(d,\mathbb{R})$ divides Ω.

(2) For $d = n(n+1)/2$, let us see $\mathbb{R}^d$ as the space $\mathrm{Sym}(n,\mathbb{R})$ of symmetric $n \times n$ real matrices, and let $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ be the image of the set of positive definite ones. The set Ω is a properly convex open subset of $X = \mathbb{P}(\mathbb{R}^d)$; it is strictly convex if and only if $n = 2$. The group $\mathrm{GL}(n,\mathbb{R})$ acts on $\mathrm{Sym}(n,\mathbb{R})$ by $g \cdot s := gsg^t$, which induces an action of $\mathrm{PGL}(n,\mathbb{R})$ on Ω. This action is transitive and the stabilizer of a point is $\mathrm{PO}(n)$, hence Ω identifies with the Riemannian symmetric space $\mathrm{PGL}(n,\mathbb{R})/\mathrm{PO}(n)$. In particular, any uniform lattice Γ of $\mathrm{PGL}(n,\mathbb{R})$ divides Ω. (A similar construction works over the complex numbers, the quaternions, or the octonions: see Benoist [ibid.].)

Many nonsymmetric strictly convex examples were also constructed since the 1960s by various techniques; see Benoist [2008] and Choi, G.-S. Lee, and Marquis [2016b] for references. Remarkably, there exist irreducible divisible convex sets $\Omega \subset \mathbb{P}(\mathbb{R}^d)$ which are not symmetric and not strictly convex: the first examples were built by Benoist [2006] for $4 \leq d \leq 7$. Ballas, Danciger, and G.-S. Lee [2018] generalized Benoist's construction for $d = 4$ to show that large families of nonhyperbolic closed 3-manifolds admit convex projective structures. Choi, G.-S. Lee, and Marquis [2016a] recently built nonstrictly convex examples of a different flavor for $5 \leq d \leq 7$.

For *strictly convex* Ω, dynamics "at infinity" are relatively well understood: if Γ divides Ω, then Γ is Gromov hyperbolic (Benoist [2004]) and, by cocompactness, any orbital map $\Gamma \to \Omega$ extends continuously to an equivariant homeomorphism ξ from the boundary

$\partial_\infty\Gamma$ of Γ to the boundary of Ω in X. This is similar to Section 3.1, except that now X itself is a flag variety G/P (see Table 1). The image of the boundary map ξ is again a *limit set* Λ_Γ on which the action of Γ is "chaotic", but Λ_Γ is now part of a larger "extended limit set" $\mathcal{L}_\Gamma$, namely the union of all projective hyperplanes tangent to Ω at points of Λ_Γ. The space $X \simeq G/P$ is the disjoint union of $\mathcal{L}_\Gamma$ and Ω. The dynamics of Γ on X are further understood by considering the *geodesic flow* on $\Omega \subset X$, defined using the Hilbert metric of Remark 3.1; for $\Omega = \mathbb{H}^n$ as in Example 3.2.(1), this is the usual geodesic flow. Benoist [ibid.] proved that the induced flow on $\Gamma\backslash\Omega$ is Anosov and topologically mixing; see Crampon [2014] for further properties.

For *nonstrictly convex* Ω, the situation is less understood. Groups Γ dividing Ω are never Gromov hyperbolic (Benoist [2004]); for $d = 4$ they are relatively hyperbolic (Benoist [2006]), but in general they might not be (e.g. if Ω is symmetric), and it is not obvious what type of boundary $\partial_\infty\Gamma$ (defined independently of Ω) might be most useful in the context of Problem B. The geodesic flow on $\Gamma\backslash\Omega$ is not Anosov, but Bray [2017] proved it is still topologically mixing for $d = 4$. Much of the dynamics remains to be explored.

By Koszul [1968], discrete subgroups of G dividing Ω are structurally stable; moreover, for a closed manifold M with fundamental group $\Gamma = \pi_1(M)$, the set $\mathrm{Hom}_M^{\mathrm{conv}}(\Gamma, G)$ of holonomy representations of convex (G, X)-structures on M is open in $\mathrm{Hom}(\Gamma, G)$. This set is also closed in $\mathrm{Hom}(\Gamma, G)$ as soon as Γ does not contain an infinite normal abelian subgroup, by Choi and Goldman [2005] (for $d = 3$) and Benoist [2005] (in general). For $d = 3$, when M is a closed surface of genus $g \geq 2$, Goldman [1990] showed that $\mathrm{Hom}_M^{\mathrm{conv}}(\Gamma, G)/G$ is homeomorphic to $\mathbb{R}^{16g-16}$, via an explicit parametrization generalizing classical (*Fenchel–Nielsen*) coordinates on Teichmüller space.

3.3 AdS quasi-Fuchsian representations. We now discuss the Lorentzian counterparts of Example 1.2, which have been studied by Witten [1988] and others as simple models for $(2+1)$-dimensional gravity. Let $M = S \times (0, 1)$ be as in Example 1.2. Instead of taking $(G, X) = (\mathrm{PO}(3, 1), \mathbb{H}^3)$, we now take $G = \mathrm{PO}(2, 2)$ and

$$X = \mathrm{AdS}^3 = \{[v] \in \mathbb{P}(\mathbb{R}^4) \mid \langle v, v\rangle_{2,2} < 0\}.$$

In other words, we change the signature of the quadratic form defining X from $(3, 1)$ (as in Example 3.2.(1)) to $(2, 2)$. This changes the natural G-invariant metric from Riemannian to Lorentzian, and the topology of X from a ball to a solid torus. The space $X = \mathrm{AdS}^3$ is called the *anti-de Sitter* 3-space.

The manifold $M = S \times (0, 1)$ does not admit (G, X)-structures of type C (terminology of Section 1.1), but it admits some of type U, called *globally hyperbolic maximal Cauchy-compact* (GHMC). In general, a Lorentzian manifold is called globally hyperbolic if it satisfies the intuitive property that "when moving towards the future one does not come back to the past"; more precisely, there is a spacelike hypersurface (*Cauchy hypersurface*) meeting each inextendible causal curve exactly once. Here we also require that the Cauchy surface be compact and that M be maximal (i.e. not isometrically embeddable into a larger globally hyperbolic Lorentzian 3-manifold).

To describe the GHMC (G, X)-structures on M, it is convenient to consider a different model for AdS^3, which leads to beautiful links with 2-dimensional hyperbolic geometry. Namely, we view $\mathbb{R}^4$ as the space $\mathfrak{M}_2(\mathbb{R})$ of real 2×2 matrices, and the quadratic form $\langle\cdot,\cdot\rangle_{2,2}$ as minus the determinant. This induces an identification of $X = \mathrm{AdS}^3$ with $\underline{G} = \mathrm{PSL}(2,\mathbb{R})$ sending $[v] \in X$ to $\left[\frac{1}{|\langle v,v\rangle|}\begin{pmatrix} v_1+v_4 & v_2+v_3 \\ v_2-v_3 & -v_1+v_4 \end{pmatrix}\right] \in \underline{G}$, and a corresponding group isomorphism from the identity component $G_0 = \mathrm{PO}(2,2)_0$ of G acting on $X = \mathrm{AdS}^3$, to $\underline{G} \times \underline{G}$ acting on $\underline{G}$ by right and left multiplication: $(g_1, g_2) \cdot g = g_2 g g_1^{-1}$. It also induces an identification of the boundary $\partial X \subset \mathbb{P}(\mathbb{R}^4)$ with the projectivization of the set of rank-one matrices, hence with $\mathbb{P}^1(\mathbb{R}) \times \mathbb{P}^1(\mathbb{R})$ (by taking the kernel and the image); the action of G_0 on ∂X corresponds to the natural action of $\underline{G} \times \underline{G}$ on $\mathbb{P}^1(\mathbb{R}) \times \mathbb{P}^1(\mathbb{R})$.

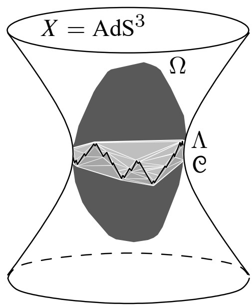


Figure 5: The sets Λ, Ω, $\mathcal{C}$ for an AdS quasi-Fuchsian representation

With these identifications, Mess [2007] proved that all GHMC (G, X)-structures on $M = S \times (0,1)$ are obtained as follows. Let (ρ_L, ρ_R) be a pair of Fuchsian representations from $\Gamma = \pi_1(M) \simeq \pi_1(S)$ to $\underline{G} = \mathrm{PSL}(2,\mathbb{R})$. The group $(\rho_L, \rho_R)(\Gamma) \subset \underline{G} \times \underline{G} \subset G$ preserves a topological circle Λ in ∂X, namely the graph of the homeomorphism of $\mathbb{P}^1(\mathbb{R})$ conjugating the action of ρ_L to that of ρ_R. For any $z \in \Lambda$, the orthogonal $z^\perp$ of z for $\langle\cdot,\cdot\rangle_{2,2}$ is a projective hyperplane tangent to X at z. The complement Ω in $\mathbb{P}(\mathbb{R}^4)$ of the union of all $z^\perp$ for $z \in \Lambda$ is a convex open subset of $\mathbb{P}(\mathbb{R}^4)$ contained in X (see Figure 5) which admits a Γ-invariant Cauchy surface. The action of Γ on Ω via (ρ_L, ρ_R) is properly discontinuous and the convex hull $\mathcal{C}$ of Λ in Ω (called the *convex core*) has compact quotient by Γ. The quotient $(\rho_L, \rho_R)(\Gamma)\backslash\Omega$ is diffeomorphic to $M = S \times (0,1)$, and this yields a GHMC (G, X)-structure on M.

Such (G, X)-structures, or their holonomy representations $\rho = (\rho_L, \rho_R) : \Gamma \to \underline{G} \times \underline{G} \subset G$, are often called *AdS quasi-Fuchsian*, by analogy with Example 1.2. Their deformation space is parametrized by $\mathrm{Teich}(S) \times \mathrm{Teich}(S)$, via (ρ_L, ρ_R) (Mess [ibid.]). Their geometry, especially the geometry of the convex core and the way it determines (ρ_L, ρ_R), is the object of active current research (see Bonsante and Schlenker [2012] and Barbot, Bonsante, Danciger, Goldman, Guéritaud, Kassel, Krasnov, Schlenker, and Zeghib [2012]). Generalizations have recently been worked out in several directions (see Bonsante, Krasnov, and Schlenker [2011], Barbot and Mérigot [2012], Barbot [2015] and Section 6.2).

As in Section 3.1, the compactness of the convex core of an AdS quasi-Fuchsian manifold implies that any orbital map $\Gamma \to \Omega$ extends "at infinity" to an equivariant embedding $\xi : \partial_\infty \Gamma \to \partial X$ with image Λ. Here ∂X is still a flag variety G/P, where P is the stabilizer in $G = \mathrm{PO}(2,2)$ of an isotropic line of $\mathbb{R}^4$ for $\langle \cdot, \cdot \rangle_{2,2}$. Although G has higher rank, the rank-one dynamics of Section 3.1 still appear through the product structure of $G_0 \simeq \underline{G} \times \underline{G}$ acting on $\partial X \simeq \mathbb{P}^1(\mathbb{R}) \times \mathbb{P}^1(\mathbb{R}) \simeq \partial_\infty \mathbb{H}^2 \times \partial_\infty \mathbb{H}^2$.

4 Anosov representations

In this section we define and discuss Anosov representations. These are representations of Gromov hyperbolic groups to Lie groups G with strong dynamical properties, defined using continuous equivariant boundary maps. They were introduced by Labourie [2006] and further investigated by Guichard and Wienhard [2012]. They play an important role in higher Teichmüller theory and in the study of Problem B. As we shall see in Section 4.5, most representations that appeared in Section 3 were in fact Anosov.

4.1 The definition. Let Γ be a Gromov hyperbolic group with boundary $\partial_\infty \Gamma$ (e.g. Γ a surface group and $\partial_\infty \Gamma$ a circle, or Γ a nonabelian free group and $\partial_\infty \Gamma$ a Cantor set). The notion of an Anosov representation of Γ to a reductive Lie group G depends on the choice of a parabolic subgroup P of G up to conjugacy, i.e. on the choice of a flag variety G/P (see Section 1.4). Here, for simplicity, we restrict to $G = \mathrm{PGL}(d, \mathbb{R})$. We choose an integer $i \in [1, d-1]$ and denote by P_i the stabilizer in G of an i-plane of $\mathbb{R}^d$, so that G/P_i identifies with the Grassmannian $\mathrm{Gr}_i(\mathbb{R}^d)$.

By definition, a representation $\rho : \Gamma \to \mathrm{PGL}(d, \mathbb{R})$ is *P_i-Anosov* if there exist two continuous ρ-equivariant maps $\xi_i : \partial_\infty \Gamma \to \mathrm{Gr}_i(\mathbb{R}^d)$ and $\xi_{d-i} : \partial_\infty \Gamma \to \mathrm{Gr}_{d-i}(\mathbb{R}^d)$ which are transverse (i.e. $\xi_i(\eta) + \xi_{d-i}(\eta') = \mathbb{R}^d$ for all $\eta \neq \eta'$ in $\partial_\infty \Gamma$) and satisfy a uniform contraction/expansion condition analogous to the one defining Anosov flows.

Let us state this condition in the original case considered by Labourie [2006], where $\Gamma = \pi_1(M)$ for some closed negatively curved manifold M. We denote by $\widetilde{M}$ the universal cover of M, by T^1 the unit tangent bundle, and by $(\varphi_t)_{t \in \mathbb{R}}$ the geodesic flow on either

$T^1(M)$ or $T^1(\widetilde{M})$. Let
$$E^\rho = \Gamma\backslash(T^1(\widetilde{M}) \times \mathbb{R}^d)$$
be the natural flat vector bundle over $T^1(M) = \Gamma\backslash T^1(\widetilde{M})$ associated to ρ, where Γ acts on $T^1(\widetilde{M})\times\mathbb{R}^d$ by $\gamma\cdot(\tilde{x}, v) = (\gamma\cdot\tilde{x}, \rho(\gamma)\cdot v)$. The geodesic flow $(\varphi_t)_{t\in\mathbb{R}}$ on $T^1(M)$ lifts to a flow $(\psi_t)_{t\in\mathbb{R}}$ on E^ρ, given by $\psi_t \cdot [(\tilde{x}, v)] = [(\varphi_t \cdot \tilde{x}, v)]$. For each $\tilde{x} \in T^1(\widetilde{M})$, the transversality of the boundary maps induces a decomposition $\mathbb{R}^d = \xi_i(\tilde{x}^+) \oplus \xi_{d-i}(\tilde{x}^-)$, where $\tilde{x}^\pm = \lim_{t\to\pm\infty} \varphi_t \cdot \tilde{x}$ are the forward and backward endpoints of the geodesic defined by $\tilde{x}$, and this defines a decomposition of the vector bundle E^ρ into the direct sum of two subbundles $E^\rho_i = \{[(\tilde{x}, v)] \mid v \in \xi_i(\tilde{x}^+)\}$ and $E^\rho_{d-i} = \{[(\tilde{x}, v)] \mid v \in \xi_{d-i}(\tilde{x}^-)\}$. This decomposition is invariant under the flow (ψ_t). By definition, the representation ρ is P_i-Anosov if the following condition is satisfied.

Condition 4.1. *The flow $(\psi_t)_{t\in\mathbb{R}}$ uniformly contracts E^ρ_i with respect to E^ρ_{d-i}, i.e. there exist $C, C' > 0$ such that for any $t \geq 0$, any $x \in T^1(M)$, and any nonzero $w_i \in E^\rho_i(x)$ and $w_{d-i} \in E^\rho_{d-i}(x)$,*
$$\frac{\|\psi_t \cdot w_i\|_{\varphi_t\cdot x}}{\|\psi_t \cdot w_{d-i}\|_{\varphi_t\cdot x}} \leq e^{-Ct+C'} \frac{\|w_i\|_x}{\|w_{d-i}\|_x},$$
where $(\|\cdot\|_x)_{x\in T^1(M)}$ is any fixed continuous family of norms on the fibers $E^\rho(x)$.

See Bridgeman, Canary, Labourie, and Sambarino [2015] for an interpretation in terms of *metric Anosov flows* (or Smale flows).

Condition 4.1 implies in particular that the boundary maps ξ_i, ξ_{d-i} are *dynamics-preserving*, in the sense that the image of the attracting fixed point in $\partial_\infty\Gamma$ of any infinite-order element $\gamma \in \Gamma$ is an attracting fixed point in $\mathrm{Gr}_i(\mathbb{R}^d)$ or $\mathrm{Gr}_{d-i}(\mathbb{R}^d)$ of $\rho(\gamma)$. By density of such fixed points in $\partial_\infty\Gamma$ and by continuity, it follows that ξ_i and ξ_{d-i} are unique, compatible (i.e. $\xi_{\min(i,d-i)}(\eta) \subset \xi_{\max(i,d-i)}(\eta)$ for all $\eta \in \partial_\infty\Gamma$), and injective (using transversality). Since $\partial_\infty\Gamma$ is compact, they are homeomorphisms onto their image.

We note that *P_i-Anosov is equivalent to P_{d-i}-Anosov*, as the integers i and $d-i$ play a similar role in the definition (up to reversing the flow, which switches contraction and expansion). In particular, we may restrict to P_i-Anosov for $1 \leq i \leq d/2$.

Guichard and Wienhard [2012] observed that an analogue of Condition 4.1 can actually be defined for any Gromov hyperbolic group Γ. The idea is to replace $T^1(\widetilde{M})$ by $\partial_\infty\Gamma^{(2)} \times \mathbb{R}$ where $\partial_\infty\Gamma^{(2)}$ is the space of pairs of distinct points in the Gromov boundary $\partial_\infty\Gamma$ of Γ, and the flow φ_t by translation by t along the $\mathbb{R}$ factor. The work of Gromov [1987] (see also Mathéus [1991], Champetier [1994], Mineyev [2005]) yields an appropriate extension of the Γ-action on $\partial_\infty\Gamma^{(2)}$ to $\partial_\infty\Gamma^{(2)} \times \mathbb{R}$, which is properly discontinuous and cocompact.

This leads to a notion of an Anosov representation for any Gromov hyperbolic group Γ, see Guichard and Wienhard [2012].

4.2 Important properties and examples. A fundamental observation motivating the study of Anosov representations is the following: if G is a semisimple Lie group of real rank one, then a representation $\rho : \Gamma \to G$ is Anosov if and only if it is convex cocompact in the sense of Section 3.1.

Moreover, many important properties of convex cocompact representations to rank-one Lie groups generalize to Anosov representations. For instance, Anosov representations $\rho : \Gamma \to G$ are quasi-isometric embeddings, see Labourie [2006] and Guichard and Wienhard [2012]; in particular, they have finite kernel and discrete image. Also by Labourie [2006] and Guichard and Wienhard [2012], any Anosov subgroup (i.e. the image of any Anosov representation $\rho : \Gamma \to G$) is structurally stable; moreover, ρ admits a neighborhood in $\mathrm{Hom}(\Gamma, G)$ consisting entirely of Anosov representations. This is due to the uniform hyperbolicity nature of the Anosov condition.

Kapovich, Leeb, and Porti, in a series of papers (see Kapovich, Leeb, and Porti [2016, 2017], Kapovich and Leeb [2017] and Guichard [2017]), have developed a detailed analogy between Anosov representations to higher-rank semisimple Lie groups and convex cocompact representations to rank-one simple groups, from the point of view of dynamics (e.g. extending the expansion property at the limit set of Section 3.1 and other classical characterizations) and topology (e.g. compactifications).

Here are some classical examples of Anosov representations in higher real rank.

Examples 4.2. Let $\Gamma = \pi_1(S)$ where S is a closed orientable surface of genus ≥ 2.

(1) (Labourie [2006]) For $d \geq 2$, let $\tau_d : \mathrm{PSL}(2,\mathbb{R}) \to G = \mathrm{PGL}(d,\mathbb{R})$ be the irreducible representation (unique up to conjugation by G). For any Fuchsian representation $\rho_0 : \Gamma \to \mathrm{PSL}(2,\mathbb{R})$, the composition $\tau_d \circ \rho_0 : \Gamma \to G$ is P_i-Anosov for all $1 \leq i \leq d-1$. Moreover, any representation in the connected component of $\tau_d \circ \rho_0$ in $\mathrm{Hom}(\Gamma, G)$ is still P_i-Anosov for all $1 \leq i \leq d-1$. These representations were first studied by Hitchin [1992] and are now known as *Hitchin representations*.

(2) (Burger, Iozzi, Labourie, and Wienhard [2005] and Burger, Iozzi, and Wienhard [2010a]) If a representation of Γ to $G = \mathrm{PSp}(2n,\mathbb{R}) \subset \mathrm{PGL}(2n,\mathbb{R})$ (resp. $G = \mathrm{PO}(2,q) \subset \mathrm{PGL}(2+q,\mathbb{R})$) is *maximal*, then it is P_n-Anosov (resp. P_1-Anosov).

(3) (Barbot [2010] for $d = 3$) Let $d \geq 2$. Any Fuchsian representation $\Gamma \to \mathrm{SL}(2,\mathbb{R})$, composed with the standard embedding $\mathrm{SL}(2,\mathbb{R}) \hookrightarrow \mathrm{SL}(d,\mathbb{R})$ (given by the direct sum of the standard action on $\mathbb{R}^2$ and the trivial action on $\mathbb{R}^{d-2}$), defines a P_1-Anosov representation $\Gamma \to G = \mathrm{PSL}(d,\mathbb{R})$.

In (2), we say that $\rho : \Gamma \to G$ is *maximal* if it maximizes a topological invariant, the *Toledo number* $T(\rho)$, defined for any simple Lie group G of Hermitian type. If $X = G/K$

is the Riemannian symmetric space of G, then the imaginary part of the G-invariant Hermitian form on X defines a real 2-form ω_X, and by definition $T(\rho) = \frac{1}{2\pi} \int_S f^* \omega_X$ where $f : \widetilde{S} \to X$ is any ρ-equivariant smooth map. For $G = \mathrm{PSL}(2, \mathbb{R})$, this coincides with the Euler number of ρ. In general, $T(\rho)$ takes discrete values and $|T(\rho)| \leq \mathrm{rank}_{\mathbb{R}}(G)\, |\chi(S)|$ where $\chi(S)$ is the Euler characteristic of S (see Burger, Iozzi, and Wienhard [2014]).

While (1) and (3) provide Anosov representations in two of the three connected components of $\mathrm{Hom}(\Gamma, \mathrm{PSL}(3, \mathbb{R}))$ for $\Gamma = \pi_1(S)$, it is currently not known whether Anosov representations appear in the third component.

See Benoist [1996], Guichard and Wienhard [2012], Kapovich, Leeb, and Porti [2014a], Canary, M. Lee, Sambarino, and Stover [2017], Burelle and Treib [2018] for higher-rank Anosov generalizations of Schottky groups.

4.3 Higher Teichmüller spaces of Anosov representations. Anosov representations play an important role in *higher Teichmüller theory*, a currently very active theory whose goal is to find deformation spaces of faithful and discrete representations of discrete groups Γ into higher-rank semisimple Lie groups G which share some of the remarkable properties of Teichmüller space. Although various groups Γ may be considered, the foundational case is when $\Gamma = \pi_1(S)$ for some closed connected surface S of genus ≥ 2 (see Burger, Iozzi, and Wienhard [2014] and Wienhard [2018]); then one can use rich features of Riemann surfaces, explicit topological considerations, and powerful techniques based on Higgs bundles as in the pioneering work of Hitchin [1992].

Strikingly similar properties to $\mathrm{Teich}(S)$ have been found for two types of *higher Teichmüller spaces*: the space of *Hitchin representations* of Γ into a real split simple Lie group G such as $\mathrm{PGL}(d, \mathbb{R})$, modulo conjugation by G; and the space of *maximal representations* of Γ into a simple Lie group G of Hermitian type such as $\mathrm{PSp}(2n, \mathbb{R})$ or $\mathrm{PO}(2, q)$, modulo conjugation by G. Both these spaces are unions of connected components of $\mathrm{Hom}(\Gamma, G)/G$, consisting entirely of Anosov representations (see Examples 4.2). Similarities of these spaces to $\mathrm{Teich}(S)$ include:

(1) the proper discontinuity of the action of the mapping class group $\mathrm{Mod}(S)$ (Wienhard [2006], Labourie [2008]);

(2) for Hitchin components: the topology of $\mathbb{R}^{\dim(G)\,|\chi(S)|}$ (Hitchin [1992]);

(3) good systems of coordinates generalizing those on $\mathrm{Teich}(S)$ (Goldman [1990], Fock and Goncharov [2006], Bonahon and Dreyer [2014], Strubel [2015], Zhang [2015]);

(4) an analytic $\mathrm{Mod}(S)$-invariant Riemannian metric, the *pressure metric* (Bridgeman, Canary, Labourie, and Sambarino [2015] and Pollicott and Sharp [2017]);

(5) versions of the collar lemma for associated locally symmetric spaces (G.-S. Lee and Zhang [2017], Burger and Pozzetti [2017]).

Other higher Teichmüller spaces of Anosov representations of $\pi_1(S)$ are also being explored, see Guichard and Wienhard [2016]. We refer to Section 5 for geometric structures associated to such spaces.

4.4 Characterizations. Various characterizations of Anosov representations have been developed in the past few years, by Labourie [2006], Guichard and Wienhard [2012], Kapovich, Leeb, and Porti [2014a,b], Kapovich and Leeb [2017], Guéritaud, Guichard, Kassel, and Wienhard [2017a], and others. Here are some characterizations that do not involve any flow. They hold for any reductive Lie group G, but for simplicity we state them for $G = \mathrm{PGL}(d,\mathbb{R})$. For $1 \leq i \leq d$ and $g \in \mathrm{GL}(d,\mathbb{R})$, we denote by $\mu_i(g)$ (resp. $\lambda_i(g)$) the logarithm of the i-th singular value (resp. eigenvalue) of g.

Theorem 4.3. *For a Gromov hyperbolic group Γ, a representation $\rho : \Gamma \to G = \mathrm{PGL}(d,\mathbb{R})$, and an integer $1 \leq i \leq d/2$, the following are equivalent:*

(1) ρ is P_i-Anosov (or equivalently P_{d-i}-Anosov, see Section 4.1);

(2) there exist continuous, ρ-equivariant, transverse, dynamics-preserving boundary maps $\xi_i : \partial_\infty\Gamma \to \mathrm{Gr}_i(\mathbb{R}^d)$ and $\xi_{d-i} : \partial_\infty\Gamma \to \mathrm{Gr}_{d-i}(\mathbb{R}^d)$, and $(\mu_i - \mu_{i+1})(\rho(\gamma)) \to +\infty$ as $|\gamma| \to +\infty$;

(3) there exist continuous, ρ-equivariant, transverse, dynamics-preserving boundary maps $\xi_i : \partial_\infty\Gamma \to \mathrm{Gr}_i(\mathbb{R}^d)$ and $\xi_{d-i} : \partial_\infty\Gamma \to \mathrm{Gr}_{d-i}(\mathbb{R}^d)$, and $(\lambda_i - \lambda_{i+1})(\rho(\gamma)) \to +\infty$ as $\ell_\Gamma(\gamma) \to +\infty$;

(4) there exist $C, C' > 0$ such that $(\mu_i - \mu_{i+1})(\rho(\gamma)) \geq C\,|\gamma| - C'$ for all $\gamma \in \Gamma$;

(5) there exist $C, C' > 0$ such that $(\lambda_i - \lambda_{i+1})(\rho(\gamma)) \geq C\,\ell_\Gamma(\gamma) - C'$ for all $\gamma \in \Gamma$.

Here we denote by $|\cdot| : \Gamma \to \mathbb{N}$ the word length with respect to some fixed finite generating subset of Γ, and by $\ell_\Gamma : \Gamma \to \mathbb{N}$ the translation length in the Cayley graph of Γ for that subset, i.e. $\ell_\Gamma(\gamma) = \min_{\beta\in\Gamma} |\beta\gamma\beta^{-1}|$. In a Gromov hyperbolic group Γ the translation length $\ell_\Gamma(\gamma)$ is known to differ only by at most a uniform additive constant from the stable length $|\gamma|_\infty = \lim_{n\to+\infty} |\gamma^n|/n$, and so we may replace $\ell_\Gamma(\gamma)$ by $|\gamma|_\infty$ in Conditions (3) and (5).

The equivalence (1) ⇔ (2) is proved in Guéritaud, Guichard, Kassel, and Wienhard [ibid.] and Kapovich, Leeb, and Porti [2014b], the equivalence (2) ⇔ (3) in Guéritaud, Guichard, Kassel, and Wienhard [2017a], the equivalence (1) ⇔ (4) in Kapovich, Leeb, and Porti [2014a] and Bochi, Potrie, and Sambarino [2018], and the equivalence (4) ⇔ (5) in Kassel and Potrie [2018].

Condition (4) is a refinement of the condition of being a quasi-isometric embedding, which for $G = \mathrm{PGL}(d,\mathbb{R})$ is equivalent to the existence of $C, C' > 0$ such that $\sqrt{\sum_k (\mu_k - \mu_{k+1})}$ $C\,|\gamma| - C'$ for all $\gamma \in \Gamma$. We refer to Guéritaud, Guichard, Kassel, and Wienhard [2017a] (*CLI condition*) and Kapovich, Leeb, and Porti [2014a] (*Morse condition*) for further refinements satisfied by Anosov representations.

By Kapovich, Leeb, and Porti [ibid.] and Bochi, Potrie, and Sambarino [2018], if Γ is any finitely generated group, then the existence of a representation $\rho : \Gamma \to \mathrm{PGL}(d,\mathbb{R})$ satisfying Condition (4) implies that Γ is Gromov hyperbolic. The analogue for (5) is more subtle: e.g. the Baumslag–Solitar group $\mathrm{BS}(1,2)$, which is not Gromov hyperbolic, still admits a faithful representation into $\mathrm{PSL}(2,\mathbb{R})$ satisfying Condition (5) for the stable length $|\cdot|_\infty$, see Kassel and Potrie [2018].

The original proof of (1) ⇔ (4) by Kapovich, Leeb, and Porti [2014a] uses the geometry of higher-rank Riemannian symmetric spaces and asymptotic cones. The alternative proof given by Bochi, Potrie, and Sambarino [2018] is based on an interpretation of (1) and (4) in terms of partially hyperbolic dynamics, and more specifically of dominated splittings for locally constant linear cocycles over certain subshifts. Pursuing this point of view further, it is shown in Kassel and Potrie [2018] that the equivalence (4) ⇔ (5) of Theorem 4.3 implies the equivalence between nonuniform hyperbolicity (i.e. all invariant measures are hyperbolic) and uniform hyperbolicity for a certain cocycle naturally associated with ρ on the space of biinfinite geodesics of Γ. In general in smooth dynamics, nonuniform hyperbolicity does not imply uniform hyperbolicity.

4.5 Revisiting the examples of Section 3. The boundary maps and dynamics "at infinity" that appeared in the examples of Section 3 are explained for the most part by the notion of an Anosov representation:

- convex cocompact representations to rank-one simple Lie groups as in Section 3.1 are all Anosov (see Section 4.2);

- if S is a closed orientable connected surface of genus ≥ 2, then by Goldman [1990] and Choi and Goldman [2005] the holonomy representations of convex projective structures on S as in Section 3.2 are exactly the Hitchin representations of $\pi_1(S)$ into $\mathrm{PSL}(3,\mathbb{R})$; they are all P_1-Anosov (Example 4.2.(1));

- for general $d \geq 3$, the work of Benoist [2004] shows that if Γ is a discrete subgroup of $\mathrm{PGL}(d,\mathbb{R})$ dividing a *strictly convex* open subset Ω of $\mathbb{P}(\mathbb{R}^d)$, then Γ is Gromov hyperbolic and the inclusion $\Gamma \hookrightarrow \mathrm{PGL}(d,\mathbb{R})$ is P_1-Anosov;

- The work of Mess [2007] implies that a representation $\rho : \pi_1(S) \to \mathrm{PO}(2,2) \subset \mathrm{PGL}(4,\mathbb{R})$ is AdS quasi-Fuchsian if and only if it is P_1-Anosov.

5 Geometric structures for Anosov representations

We just saw in Section 4.5 that various (G, X)-structures described in Section 3 give rise (via the holonomy) to Anosov representations; these (G, X)-structures are of type C or type U (terminology of Section 1.1). In this section, we study the converse direction. Namely, given an Anosov representation $\rho : \Gamma \to G$, we wish to find:

- homogeneous spaces $X = G/H$ on which Γ acts properly discontinuously via ρ; this will yield (G, X)-manifolds (or orbifolds) $M = \rho(\Gamma)\backslash X$ of type C;
- proper open subsets $\mathcal{U}$ (*domains of discontinuity*) of homogeneous spaces $X = G/H$ on which Γ acts properly discontinuously via ρ; this will yield (G, X)-manifolds (or orbifolds) $M = \rho(\Gamma)\backslash\mathcal{U}$ of type U.

We discuss type U in Sections 5.1 and 5.2 and type C in Section 5.3. One motivation is to give a geometric meaning to the higher Teichmüller spaces of Section 4.3.

5.1 Cocompact domains of discontinuity. Domains of discontinuity with compact quotient have been constructed in several settings in the past ten years.

Barbot [2010] constructed such domains in the space X of flags of $\mathbb{R}^3$, for the Anosov representations to $G = \mathrm{PSL}(3,\mathbb{R})$ of Example 4.2.(3) and their small deformations.

Guichard and Wienhard [2012] developed a more general construction of cocompact domains of discontinuity in flag varieties X for Anosov representations to semisimple Lie groups G. Here is one of their main results. For $p \geq q \geq i \geq 1$, we denote by $\mathfrak{F}_i^{p,q}$ the closed subspace of the Grassmannian $\mathrm{Gr}_i(\mathbb{R}^{p+q})$ consisting of i-planes that are totally isotropic for the standard symmetric bilinear form $\langle\cdot,\cdot\rangle_{p,q}$ of signature (p, q).

Theorem 5.1 (Guichard and Wienhard [ibid.]). *Let $G = \mathrm{PO}(p,q)$ with $p \geq q$ and $X = \mathfrak{F}_q^{p,q}$. For any P_1-Anosov representation $\rho : \Gamma \to G \subset \mathrm{PGL}(\mathbb{R}^{p+q})$ with boundary map $\xi_1 : \partial_\infty\Gamma \to \mathfrak{F}_1^{p,q} \subset \mathbb{P}(\mathbb{R}^{p+q})$, the group $\rho(\Gamma)$ acts properly discontinuously with compact quotient on $\mathcal{U}_\rho := X \smallsetminus \mathcal{L}_\rho$, where*

$$\mathcal{L}_\rho := \bigcup_{\eta\in\partial_\infty\Gamma} \{W \in X = \mathfrak{F}_q^{p,q} \mid \xi_1(\eta) \in W\}.$$

We have $\mathcal{U}_\rho \neq \varnothing$ as soon as $\dim(\partial_\infty\Gamma) < p-1$. The homeomorphism type of $\rho(\Gamma)\backslash\mathcal{U}_\rho$ is constant as ρ varies continuously among P_1-Anosov representations of Γ to G.

For $q = 1$, we recover the familiar picture of Section 3.1: the set $\mathcal{L}_\rho$ is the limit set $\Lambda_{\rho(\Gamma)} \subset \partial_\infty \mathbb{H}^p$, and $\mathcal{U}_\rho$ is the domain of discontinuity $\Omega_{\rho(\Gamma)} = \partial_\infty \mathbb{H}^p \smallsetminus \Lambda_{\rho(\Gamma)}$.

For $q = 2$, Theorem 5.1 fits into the theory of *Lorentzian Kleinian groups* acting on the Einstein universe $X = \mathrm{Ein}^p = \mathfrak{F}_1^{p,2}$, as developed by Frances [2005].

Guichard and Wienhard [2012] used Theorem 5.1 to describe domains of discontinuity for various families of Anosov representations to other semisimple Lie groups G. Indeed, they proved that an Anosov representation $\rho : \Gamma \to G$ can always be composed with a representation of G to some $\mathrm{PO}(p,q)$ so as to become P_1-Anosov in $\mathrm{PO}(p,q)$.

Kapovich, Leeb, and Porti [2018] developed a more systematic approach to the construction of domains of discontinuity in flag varieties. They provided sufficient conditions (expressed in terms of a notion of *balanced* ideal in the Weyl group for the Bruhat order) on triples (G, P, Q) consisting of a semisimple Lie group G and two parabolic subgroups P and Q, so that P-Anosov representations to G admit cocompact domains of discontinuity in G/Q. These domains are obtained by removing an explicit "extended limit set" $\mathcal{L}_\rho$ as in Theorem 5.1. The approach of Kapovich–Leeb–Porti is intrinsic: it does not rely on an embedding of G into some $\mathrm{PO}(p,q)$.

5.2 Geometric structures for Hitchin and maximal representations. Let $\Gamma = \pi_1(S)$ where S is a closed orientable surface of genus ≥ 2. Recall (Examples 4.2) that Hitchin representations from Γ to $G = \mathrm{PSL}(d,\mathbb{R})$ are P_i-Anosov for all $1 \leq i \leq d-1$; maximal representations from Γ to $G = \mathrm{PO}(2,q) \subset \mathrm{PGL}(2+q,\mathbb{R})$ are P_1-Anosov.

For $G = \mathrm{PSL}(2,\mathbb{R}) \simeq \mathrm{PO}(2,1)_0$, Hitchin representations and maximal representations of Γ to G both coincide with the Fuchsian representations; they are the holonomy representations of hyperbolic structures on S (Example 2.1). In the setting of higher Teichmüller theory (see Section 4.3), one could hope that Hitchin or maximal representations of Γ to higher-rank Lie groups G might also parametrize certain geometric structures on a manifold related to S. We saw in Section 4.5 that this is indeed the case for $G = \mathrm{PSL}(3,\mathbb{R})$: Hitchin representations of Γ to $\mathrm{PSL}(3,\mathbb{R})$ parametrize the convex projective structures on S (Goldman [1990], Choi and Goldman [2005]). In an attempt to generalize this picture, we now outline constructions of domains of discontinuity for Hitchin representations to $\mathrm{PSL}(d,\mathbb{R})$ with $d > 3$, and maximal representations to $\mathrm{PO}(2,q)$ with $q > 1$. By classical considerations of cohomological dimension, such domains cannot be both cocompact and contractible; in Sections 5.2.2 and 5.2.3 below, we will prefer to forgo compactness to favor the nice geometry of convex domains.

5.2.1 Hitchin representations for even $d = 2n$. Let $(G, X) = (\mathrm{PSL}(2n,\mathbb{R}), \mathbb{P}(\mathbb{R}^{2n}))$. Hitchin representations into G do not preserve any properly convex open set Ω in X (see Danciger, Guéritaud, and Kassel [2017], Zimmer [2017]). However, Guichard and Wienhard [2008, 2012] associated to them nonconvex (G, X)-structures on a closed manifold: if $\rho : \Gamma \to G$ is Hitchin with boundary map $\xi_n : \partial_\infty \Gamma \to \mathrm{Gr}_n(\mathbb{R}^{2n})$, then $\mathcal{U} = X \smallsetminus \bigcup_{\eta \in \partial_\infty \Gamma} \xi_n(\eta)$ is a cocompact domain of discontinuity for ρ. For $n = 2$ the quotient has two connected components which both fiber in circles over S; considering one of them, the Hitchin representations parametrize projective structures on $T^1(S)$ with a natural foliation by 2-dimensional convex sets (Guichard and Wienhard [2008]). For $3 \le n \le 63$, Alessandrini and Li [2018] recently used Higgs bundle techniques to describe a fibration of $\rho(\Gamma)\backslash\mathcal{U}$ over S with fiber $\mathrm{O}(n)/\mathrm{O}(n-2)$.

5.2.2 Hitchin representations for odd $d = 2n + 1$. Hitchin representations into $G = \mathrm{PSL}(2n + 1, \mathbb{R})$ give rise to (G, X)-manifolds for at least two choices of X.

One choice for X is the space of partial flags $(V_1 \subset V_{2n})$ of $\mathbb{R}^{2n+1}$ with V_1 a line and V_{2n} a hyperplane: Guichard and Wienhard [2012] again constructed explicit cocompact domains of discontinuity in X in this setting.

Another choice is $X = \mathbb{P}(\mathbb{R}^{2n+1})$: Hitchin representations in odd dimension are the holonomies of convex projective manifolds, which are noncompact for $n > 1$.

Theorem 5.2 (Danciger, Guéritaud, and Kassel [2017], Zimmer [2017])**.** *For any Hitchin representation $\rho : \Gamma \to \mathrm{PSL}(2n + 1, \mathbb{R})$, there is a $\rho(\Gamma)$-invariant properly convex open subset Ω of $\mathbb{P}(\mathbb{R}^{2n+1})$ and a nonempty closed convex subset $\mathcal{C}$ of Ω which has compact quotient by $\rho(\Gamma)$.*

More precisely, if ρ has boundary maps $\xi_1 : \partial_\infty \Gamma \to \mathrm{Gr}_1(\mathbb{R}^{2n+1}) = \mathbb{P}(\mathbb{R}^{2n+1})$ and $\xi_{2n} : \partial_\infty \Gamma \to \mathrm{Gr}_{2n}(\mathbb{R}^{2n+1})$, we may take $\Omega = \mathbb{P}(\mathbb{R}^{2n+1}) \smallsetminus \bigcup_{\eta \in \partial_\infty \Gamma} \xi_{2n}(\eta)$ and $\mathcal{C}$ to be the convex hull of $\xi_1(\partial_\infty \Gamma)$ in Ω. The group $\rho(\Gamma)$ acts properly discontinuously on Ω (Remark 3.1), and so $\rho(\Gamma)\backslash\Omega$ is a convex projective manifold, with a compact convex core $\rho(\Gamma)\backslash\mathcal{C}$. In other words, $\rho(\Gamma)$ is *convex cocompact in* $\mathbb{P}(\mathbb{R}^{2n+1})$, see Section 6.

5.2.3 Maximal representations. Maximal representations into $G = \mathrm{PO}(2, q)$ give rise to (G, X)-manifolds for at least two choices of X.

One choice is $X = \mathcal{F}_2^{2,q}$ (also known as the space of *photons* in the Einstein universe Ein^q): Theorem 5.1 provides cocompact domains of discontinuity for ρ in X. Collier, Tholozan, and Toulisse [2017] recently studied the geometry of the associated quotient (G, X)-manifolds, and showed that they fiber over S with fiber $\mathrm{O}(q)/\mathrm{O}(q-2)$.

Another choice is $X = \mathbb{P}(\mathbb{R}^{2+q})$: by Danciger, Guéritaud, and Kassel [2018a], maximal representations $\rho : \Gamma \to G$ are the holonomy representations of convex projective manifolds $\rho(\Gamma)\backslash\Omega$, which are noncompact for $q > 1$ but still convex cocompact as in Section 5.2.2. In fact Ω can be taken inside $\mathbb{H}^{2,q-1} = \{[v] \in \mathbb{P}(\mathbb{R}^{2+q}) \mid \langle v, v \rangle_{2,q} < 0\}$ (see also Collier, Tholozan, and Toulisse [2017]), which is a pseudo-Riemannian analogue of the real hyperbolic space in signature $(2, q-1)$, and $\rho(\Gamma)$ is $\mathbb{H}^{2,q-1}$-convex cocompact in the sense of Section 6.2 below.

5.3 Proper actions on full homogeneous spaces. In Sections 5.1 and 5.2, we mainly considered *compact* homogeneous spaces $X = G/H$ (flag varieties); these spaces cannot admit proper actions by infinite discrete groups, but we saw that sometimes they can contain domains of discontinuity $\mathcal{U} \subsetneq X$, yielding (G, X)-manifolds of type U (terminology of Section 1.1).

We now consider *noncompact* homogeneous spaces $X = G/H$. Then Anosov representations $\rho : \Gamma \to G$ may give proper actions of Γ on the whole of $X = G/H$, yielding (G, X)-manifolds $\rho(\Gamma)\backslash X$ of type C. When H is compact, this is not very interesting since all faithful and discrete representations to G give proper actions on X. However, when H is noncompact, it may be remarkably difficult in general to find such representations giving proper actions on X, which led to a rich literature (see Kobayashi and Yoshino [2005] and Kassel [2009, Intro]).

One construction for proper actions on X was initiated by Guichard and Wienhard [2012] and developed further in Guéritaud, Guichard, Kassel, and Wienhard [2017b]. Starting from an Anosov representation $\rho : \Gamma \to G$, the idea is to embed G into some larger semisimple Lie group G' so that $X = G/H$ identifies with a G-orbit in some flag variety $\mathcal{F}'$ of G', and then to find a cocompact domain of discontinuity $\mathcal{U} \supset X$ for ρ in $\mathcal{F}'$ by using a variant of Theorem 5.1. The action of $\rho(\Gamma)$ on X is then properly discontinuous, and $\rho(\Gamma)\backslash(\mathcal{U} \cap \overline{X})$ provides a compactification of $\rho(\Gamma)\backslash X$, which in many cases can be shown to be well-behaved. Here is one of the applications of this construction given in Guéritaud, Guichard, Kassel, and Wienhard [ibid.].

Example 5.3. Let $G = \mathrm{PO}(p, q)$ and $H = \mathrm{O}(p, q-1)$ where $p > q \geq 1$. For any P_q-Anosov representation $\rho : \Gamma \to G \subset \mathrm{PGL}(p+q, \mathbb{R})$, the group $\rho(\Gamma)$ acts properly discontinuously on $X = \mathbb{H}^{p,q-1} = \{[v] \in \mathbb{P}(\mathbb{R}^{p+q}) \mid \langle v, v\rangle_{p,q} < 0\} \simeq G/H$, and for torsion-free Γ the complete (G, X)-manifold $\rho(\Gamma)\backslash X$ is topologically tame.

By *topologically tame* we mean homeomorphic to the interior of a compact manifold with boundary. For other compactifications of quotients of homogeneous spaces by Anosov representations, yielding topological tameness, see Guichard, Kassel, and Wienhard [2015], Kapovich and Leeb [2015], and Kapovich, Leeb, and Porti [2018].

Another construction of complete (G, X)-manifolds for Anosov representations to reductive Lie groups G was given in Guéritaud, Guichard, Kassel, and Wienhard [2017a], based on a properness criterion of Benoist [1996] and Kobayashi [1996]. For simplicity we discuss it for $G = \mathrm{PGL}(d, \mathbb{R})$. As in Section 4.4, let $\mu_i(g)$ be the logarithm of the i-th singular value of a matrix $g \in \mathrm{GL}(d, \mathbb{R})$; this defines a map $\mu = (\mu_1, \dots, \mu_d) : \mathrm{PGL}(d, \mathbb{R}) \to \mathbb{R}^d/\mathbb{R}(1, \dots, 1) \simeq \mathbb{R}^{d-1}$. The properness criterion of Benoist and Kobayashi states that for two closed subgroups H, Γ of $G = \mathrm{PGL}(d, \mathbb{R})$, the action of Γ on G/H is properly discontinuous if and only if the set $\mu(\Gamma)$ "drifts away at infinity from $\mu(H)$", in the sense that for any $R > 0$ we have $d_{\mathbb{R}^{d-1}}(\mu(\gamma), \mu(H)) \geq R$ for all but finitely many $\gamma \in \Gamma$. If Γ is the image of an Anosov representation, then we can apply the implication (1) $\Rightarrow$ (2) of Theorem 4.3 to see that the properness criterion is satisfied for many examples of H.

Example 5.4. For $i = 1$ (resp. n), the image of any P_i-Anosov representation to $G = \mathrm{PSL}(2n, \mathbb{R})$ acts properly discontinuously on $X = G/H$ for $H = \mathrm{SL}(n, \mathbb{C})$ (resp. $\mathrm{SO}(n+1, n-1)$).

6 Convex cocompact projective structures

In Sections 3 and 4.5 we started from (G, X)-structures to produce Anosov representations, and in Section 5 we started from Anosov representations to produce (G, X)-structures. We now discuss a situation, in the setting of convex projective geometry, in which the links between (G, X)-structures and Anosov representations are particularly tight and go in both directions, yielding a better understanding of both sides. In Section 6.4 we will also encounter generalizations of Anosov representations, for finitely generated groups that are not necessarily Gromov hyperbolic.

6.1 Convex cocompactness in higher real rank. The results presented here are part of a quest to generalize the notion of rank-one convex cocompactness of Section 3.1 to higher real rank.

The most natural generalization, in the setting of Riemannian symmetric spaces, turns out to be rather restrictive: Kleiner and Leeb [2006] and Quint [2005] proved that if G is a real simple Lie group of real rank ≥ 2 and K a maximal compact subgroup of G, then any Zariski-dense discrete subgroup of G acting with compact quotient on some nonempty convex subset of G/K is a uniform lattice in G.

Meanwhile, we have seen in Section 4.2 that Anosov representations to higher-rank semisimple Lie groups G have strong dynamical properties which nicely generalize those of rank-one convex cocompact representations (see Kapovich, Leeb, and Porti [2016, 2017], Kapovich and Leeb [2017] and Guichard [2017]). However, in general Anosov representations to G do not act with compact quotient on any nonempty convex subset of G/K, and it is not clear that Anosov representations should come with any geometric notion of convexity at all (see e.g. Section 5.2.1).

In this section, we shall see that Anosov representations in fact do come with convex structures. We shall introduce several generalizations of convex cocompactness to higher real rank (which we glimpsed in Sections 5.2.2 and 5.2.3) and relate them to Anosov representations.

6.2 Convex cocompactness in pseudo-Riemannian hyperbolic spaces. We start with a generalization of the hyperbolic quasi-Fuchsian manifolds of Example 1.2 or the AdS quasi-Fuchsian manifolds of Section 3.3, where we replace the real hyperbolic space $\mathbb{H}^3$ or its Lorentzian analogue AdS^3 by their general pseudo-Riemannian analogue in signature $(p, q-1)$ for $p, q \geq 1$, namely

$$X = \mathbb{H}^{p,q-1} = \big\{[v] \in \mathbb{P}(\mathbb{R}^{p+q}) \mid \langle v, v\rangle_{p,q} < 0\big\}.$$

The symmetric bilinear form $\langle\cdot,\cdot\rangle_{p,q}$ of signature (p,q) induces a pseudo-Riemannian structure of signature $(p, q-1)$ on X, with isometry group $G = \mathrm{PO}(p,q)$ and constant negative sectional curvature (see e.g. Danciger, Guéritaud, and Kassel [2018a, § 2.1]). The following is not our original definition, but an equivalent one from Danciger, Guéritaud, and Kassel [2017, Th. 1.25].

Definition 6.1. A discrete subgroup Γ of $G = \mathrm{PO}(p,q)$ is $\mathbb{H}^{p,q-1}$*-convex cocompact* if it preserves a properly convex open subset Ω of $X = \mathbb{H}^{p,q-1} \subset \mathbb{P}(\mathbb{R}^{p+q})$ and if it acts with compact quotient on some closed convex subset $\mathcal{C}$ of Ω with nonempty interior, whose ideal boundary $\partial_{\mathrm{i}}\mathcal{C} := \overline{\mathcal{C}} \smallsetminus \mathcal{C} = \overline{\mathcal{C}} \cap \partial X$ does not contain any nontrivial projective segment. A representation $\rho : \Gamma \to G$ is $\mathbb{H}^{p,q-1}$*-convex cocompact* if its kernel is finite and its image is an $\mathbb{H}^{p,q-1}$-convex cocompact subgroup of G.

Here $\overline{\mathcal{C}}$ is the closure of $\mathcal{C}$ in $\mathbb{P}(\mathbb{R}^{p+q})$ and ∂X the boundary of $X = \mathbb{H}^{p,q-1}$ in $\mathbb{P}(\mathbb{R}^{p+q})$. For $\Gamma, \Omega, \mathcal{C}$ as in Definition 6.1, the quotient $\Gamma\backslash\Omega$ is a (G, X)-manifold (or orbifold) (see Remark 3.1), which we shall call *convex cocompact*; the subset $\Gamma\backslash\mathcal{C}$ is compact, convex, and contains all the topology, as in Sections 3.1 and 3.3.

There is a rich world of examples of convex cocompact (G, X)-manifolds, including direct generalizations of the quasi-Fuchsian manifolds of Sections 3.1 and 3.3 (see Barbot and Mérigot [2012] and Danciger, Guéritaud, and Kassel [2018a, 2017]) but also more exotic examples where the fundamental group is not necessarily realizable as a discrete subgroup of $\mathrm{PO}(p,1)$ (see Danciger, Guéritaud, and Kassel [2018a] and G.-S. Lee and Marquis [2018]).

The following result provides links with Anosov representations.

Theorem 6.2 (Danciger, Guéritaud, and Kassel [2018a, 2017])**.** *For $p, q \geq 1$, let Γ be an infinite discrete group and $\rho : \Gamma \to G = \mathrm{PO}(p,q) \subset \mathrm{PGL}(p+q, \mathbb{R})$ a representation.*

(1) If ρ is $\mathbb{H}^{p,q-1}$-convex cocompact, then Γ is Gromov hyperbolic and ρ is P_1-Anosov.

(2) *Conversely, if* Γ *is Gromov hyperbolic, if* ρ *is* P_1*-Anosov, and if* $\rho(\Gamma)$ *preserves a properly convex open subset of* $\mathbb{P}(\mathbb{R}^{p+q})$*, then* ρ *is* $\mathbb{H}^{p,q-1}$*-convex cocompact or* $\mathbb{H}^{q,p-1}$*-convex cocompact.*

(3) *If* Γ *is Gromov hyperbolic with* connected *boundary* $\partial_\infty\Gamma$ *and if* ρ *is* P_1*-Anosov, then* ρ *is* $\mathbb{H}^{p,q-1}$*-convex cocompact or* $\mathbb{H}^{q,p-1}$*-convex cocompact.*

In (2)–(3), the phrase "$\mathbb{H}^{q,p-1}$-convex cocompact" is understood after identifying $\mathrm{PO}(p,q)$ with $\mathrm{PO}(q,p)$ and $\mathbb{P}(\mathbb{R}^{p,q}) \smallsetminus \overline{\mathbb{H}^{p,q-1}}$ with $\mathbb{H}^{q,p-1}$ under multiplication of $\langle\cdot,\cdot\rangle_{p,q}$ by -1. The case that $q = 2$ and Γ is isomorphic to a uniform lattice of $\mathrm{PO}(p,1)$ is due to Barbot and Mérigot [2012].

The links between $\mathbb{H}^{p,q-1}$-convex cocompactness and Anosov representations in Theorem 6.2 have several applications.

Applications to (G, X)-structures, see Danciger, Guéritaud, and Kassel [2018a, 2017].
• $\mathbb{H}^{p,q-1}$-convex cocompactness is stable under small deformations, because being Anosov is; thus the set of holonomy representations of convex cocompact (G, X)-structures on a given manifold M is open in $\mathrm{Hom}(\pi_1(M), G)$.
• Examples of convex cocompact (G, X)-manifolds can be obtained using classical families of Anosov representations: e.g. Hitchin representations into $\mathrm{PO}(n+1,n)$ are $\mathbb{H}^{n+1,n-1}$-convex cocompact for odd n and $\mathbb{H}^{n,n}$-convex cocompact for even n, and Hitchin representations into $\mathrm{PO}(n+1,n+1)$ are $\mathbb{H}^{n+1,n}$-convex cocompact. Maximal representations into $\mathrm{PO}(2,q)$ are $\mathbb{H}^{2,q-1}$-convex cocompact, see Section 5.2.3.

Applications to Anosov representations:
• New examples of Anosov representations can be constructed from convex cocompact (G, X)-manifolds: e.g. this approach is used in Danciger, Guéritaud, and Kassel [2018a] to prove that any Gromov hyperbolic right-angled Coxeter group in d generators admits P_1-Anosov representations to $\mathrm{PGL}(d,\mathbb{R})$. This provides a large new class of hyperbolic groups admitting Anosov representations; these groups can have arbitrary large cohomological dimension, and exotic boundaries (see Dani [2017] for references). (Until now most known examples of Anosov representations were for surface groups or free groups.)
• For $q = 2$ and Γ a uniform lattice of $\mathrm{PO}(p,1)_0$, Barbot [2015] used convex cocompact (G, X)-structures to prove that the connected component $\mathcal{T}$ of $\mathrm{Hom}(\Gamma, \mathrm{PO}(p,2))$ containing the natural inclusion $\Gamma \hookrightarrow \mathrm{PO}(p,1)_0 \hookrightarrow \mathrm{PO}(p,2)$ consists entirely of Anosov representations. This is interesting in the framework of Section 4.3.

6.3 Strong projective convex cocompactness. We now consider a broader notion of convex cocompactness, not involving any quadratic form. Let $d \geq 2$.

Definition 6.3. A discrete subgroup Γ of $G = \mathrm{PGL}(d,\mathbb{R})$ is *strongly* $\mathbb{P}(\mathbb{R}^d)$*-convex cocompact* if it preserves a strictly convex open subset Ω of $X = \mathbb{P}(\mathbb{R}^d)$ with C^1 boundary and if it acts with compact quotient on some nonempty closed convex subset $\mathcal{C}$ of Ω. A representation $\rho : \Gamma \to G$ is *strongly* $\mathbb{P}(\mathbb{R}^d)$*-convex cocompact* if its kernel is finite and its image is a strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact subgroup of G.

The action of Γ on Ω in Definition 6.3 is a special case of a class of *geometrically finite actions* introduced by Crampon and Marquis [2014]. We use the adverb "strongly" to emphasize the strong regularity assumptions made on Ω. In Definition 6.3 we say that the quotient $\Gamma\backslash\Omega$ is a *strongly convex cocompact* projective manifold (or orbifold); the subset $\Gamma\backslash\mathcal{C}$ is again compact, convex, and contains all the topology.

Strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representations include $\mathbb{H}^{p,q-1}$-convex cocompact representations as in Section 6.2 (see Danciger, Guéritaud, and Kassel [2018a]), and the natural inclusion of groups dividing strictly convex open subsets of $\mathbb{P}(\mathbb{R}^d)$ as in Section 3.2. The following result generalizes Theorem 6.2, and improves on earlier results of Benoist [2004] and Crampon and Marquis [2014].

Theorem 6.4 (Danciger, Guéritaud, and Kassel [2017]). *Let Γ be an infinite discrete group and $\rho : \Gamma \to G = \mathrm{PGL}(d,\mathbb{R})$ a representation such that $\rho(\Gamma)$ preserves a nonempty properly convex open subset of $X = \mathbb{P}(\mathbb{R}^d)$. Then ρ is strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact if and only if Γ is Gromov hyperbolic and ρ is P_1-Anosov.*

Another generalization of Theorem 6.2 was independently obtained by Zimmer [2017]: it is similar to Theorem 6.4, but involves a slightly different notion of convex cocompactness and assumes $\rho(\Gamma)$ to act irreducibly on $\mathbb{P}(\mathbb{R}^d)$.

Applications of Theorem 6.4 include:

• Examples of strongly convex cocompact projective manifolds using classical Anosov representations (e.g. Hitchin representations into $\mathrm{PSL}(2n+1,\mathbb{R})$ as in Section 5.2.2).

• In certain cases, a better understanding of the set of Anosov representations of a Gromov hyperbolic group Γ inside a given connected component of $\mathrm{Hom}(\Gamma, G)$: e.g. for an irreducible hyperbolic right-angled Coxeter group Γ on k generators, it is proved in Danciger, Guéritaud, and Kassel [2018c], using Theorem 6.4 and the work of Vinberg [1971], that P_1-Anosov representations form the full interior of the space of faithful and discrete representations of Γ as a reflection group in $G = \mathrm{PGL}(d,\mathbb{R})$ when $d \geq k$.

For a Gromov hyperbolic group Γ and a P_1-Anosov representation $\rho : \Gamma \to G = \mathrm{PGL}(d,\mathbb{R})$, the group $\rho(\Gamma)$ does not always preserve a properly convex open subset of $X = \mathbb{P}(\mathbb{R}^d)$: see Section 5.2.1. However, as observed by Zimmer [2017], ρ can always be composed with the embedding $\iota : G \hookrightarrow \mathrm{PGL}(V)$ described in Example 3.2.(2), for $V = \mathrm{Sym}(d,\mathbb{R}) \simeq \mathbb{R}^{d(d+1)}$; then $\iota \circ \rho(\Gamma)$ preserves a properly convex open subset in $\mathbb{P}(V)$. The composition $\iota\circ\rho : \Gamma \to \mathrm{PGL}(V)$ is still P_1-Anosov by Guichard and Wienhard [2012], and it is strongly $\mathbb{P}(V)$-convex cocompact by Theorem 6.4. More generally, using Guichard and Wienhard [ibid.], any Anosov representation to any semisimple Lie group can always be composed with an embedding into some $\mathrm{PGL}(V)$ so as to become strongly $\mathbb{P}(V)$-convex cocompact.

6.4 Projective convex cocompactness in general. We now introduce an even broader notion of convex cocompactness, where we remove the strong regularity assumptions on Ω in Definition 6.3. This yields a large class of convex projective manifolds, whose fundamental groups are not necessarily Gromov hyperbolic. Their holonomy representations are generalizations of Anosov representations, sharing some of their desirable properties (Theorem 6.7). This shows that Anosov representations are not the only way to successfully generalize rank-one convex cocompactness to higher real rank.

Definition 6.5 (Danciger, Guéritaud, and Kassel [2017]). A discrete subgroup Γ of $G = \mathrm{PGL}(d, \mathbb{R})$ is $\mathbb{P}(\mathbb{R}^d)$-*convex cocompact* if it preserves a properly convex open subset Ω of $X = \mathbb{P}(\mathbb{R}^d)$ and if it acts with compact quotient on some "large enough" closed convex subset $\mathcal{C}$ of Ω. A representation $\rho : \Gamma \to G$ is $\mathbb{P}(\mathbb{R}^d)$-*convex cocompact* if its kernel is finite and its image is a $\mathbb{P}(\mathbb{R}^d)$-convex cocompact subgroup of G.

In Definition 6.5, by "$\mathcal{C}$ large enough" we mean that all accumulation points of all Γ-orbits of Ω are contained in the boundary of $\mathcal{C}$ in $X = \mathbb{P}(\mathbb{R}^d)$. If we did not impose this (even if we asked $\mathcal{C}$ to have nonempty interior), then the notion of $\mathbb{P}(\mathbb{R}^d)$-convex cocompactness would not be stable under small deformations: see Danciger, Guéritaud, and Kassel [2017, 2018b]. In Definition 6.5 we call $\Gamma\backslash\Omega$ a *convex cocompact* projective manifold (or orbifold).

The class of $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representations includes all strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representations as in Section 6.3, hence all $\mathbb{H}^{p,q-1}$-convex cocompact representations as in Section 6.2. In fact, the following holds.

Proposition 6.6 (Danciger, Guéritaud, and Kassel [2017]). *Let Γ be an infinite discrete group. A representation $\rho : \Gamma \to G = \mathrm{PGL}(d, \mathbb{R})$ is strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact (Definition 6.3) if and only if it is $\mathbb{P}(\mathbb{R}^d)$-convex cocompact (Definition 6.5) and Γ is Gromov hyperbolic.*

This generalizes a result of Benoist [2004] on divisible convex sets. Together with Theorem 6.4, Proposition 6.6 shows that $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representations are generalizations of Anosov representations, to a larger class of finitely generated groups Γ which are not necessarily Gromov hyperbolic. These representations still enjoy the following good properties.

Theorem 6.7 (Danciger, Guéritaud, and Kassel [2017]). *Let Γ be an infinite discrete group and $\rho : \Gamma \to G = \mathrm{PGL}(d, \mathbb{R})$ a $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representation. Then*

(1) ρ is a quasi-isometric embedding;

(2) *there is a neighborhood of* ρ *in* $\mathrm{Hom}(\Gamma, G)$ *consisting entirely of* $\mathbb{P}(\mathbb{R}^d)$*-convex cocompact representations;*

(3) ρ *is* $\mathbb{P}((\mathbb{R}^d)^*)$*-convex cocompact;*

(4) ρ *induces a* $\mathbb{P}(\mathbb{R}^D)$*-convex cocompact representation for any* $D \geq d$ *(by lifting* ρ *to a representation to* $\mathrm{SL}^{\pm}(d, \mathbb{R})$ *and composing it with the natural inclusion* $\mathrm{SL}^{\pm}(d, \mathbb{R}) \hookrightarrow \mathrm{SL}^{\pm}(D, \mathbb{R})$*).*

In order to prove (2), we show that the representations of Theorem 6.7 are exactly the holonomy representations of compact convex projective manifolds with strictly convex boundary; this allows to apply the deformation theory of Cooper, Long, and Tillmann [2018].

Groups that are $\mathbb{P}(\mathbb{R}^d)$-convex cocompact but not strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact include all groups dividing a properly convex, but not strictly convex, open subset of $X = \mathbb{P}(\mathbb{R}^d)$ as in Section 3.2, as well as their small deformations in $\mathrm{PGL}(D, \mathbb{R})$ for $D \geq d$ (Theorem 6.7.(2)–(4)). Such nontrivial deformations exist: e.g. for $d = 4$ we can always bend along tori or Klein bottle subgroups, see Benoist [2006]. There seems to be a rich world of examples beyond this, which is just starting to be unveiled, see Danciger, Guéritaud, and Kassel [2017, 2018b,c]. It would be interesting to understand the precise nature of the corresponding abstract groups Γ, and how the dynamics of $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representations generalize that of Anosov representations.

7 Complete affine structures

In Sections 3 to 6 we always considered semisimple, or more generally reductive, Lie groups G. We now discuss links between (G, X)-structures and representations of discrete groups into G in an important case where G is not reductive: namely G is the group $\mathrm{Aff}(\mathbb{R}^d) = \mathrm{GL}(d, \mathbb{R}) \ltimes \mathbb{R}^d$ of invertible affine transformations of $X = \mathbb{R}^d$. We shall see in Section 7.3 that for $d = 3$ the holonomy representations of certain complete (i.e. type C in Section 1.1) (G, X)-structures are characterized by a uniform contraction condition, which is also an *affine Anosov* condition; we shall briefly mention partial extensions to $d > 3$, which are currently being explored.

7.1 Brief overview: understanding complete affine manifolds. Let $(G, X) = (\mathrm{Aff}(\mathbb{R}^d), \mathbb{R}^d)$. This section is centered around *complete affine manifolds*, i.e. (G, X)-manifolds of the form $M = \rho(\Gamma)\backslash X$ where $\Gamma \simeq \pi_1(M)$ is a discrete group and $\rho : \Gamma \to G$ a faithful representation through which Γ acts properly discontinuously and freely on $X = \mathbb{R}^d$. The study of such representations has a rich history through the interpretation of their images as *affine crystallographic groups*, i.e. symmetry groups of affine tilings of $\mathbb{R}^d$, possibly with noncompact tiles; see Abels [2001] for a detailed survey. The compact and noncompact cases are quite different.

For a compact complete affine manifold M, Auslander [1964] conjectured that $\pi_1(M)$ must always be virtually (i.e. up to finite index) polycyclic. This extends a classical theorem of Bieberbach on affine Euclidean isometries. The conjecture is proved for $d \leq 6$ (Fried and Goldman [1983], Abels, Margulis, and Soifer [2012]), but remains wide open for $d \geq 7$, despite partial results (see Abels [2001]).

In contrast, in answer to a question of Milnor [1977], there exist noncompact complete affine manifolds M for which $\pi_1(M)$ is *not* virtually polycyclic. The first examples were constructed by Margulis [1984] for $d = 3$, with $\pi_1(M)$ a nonabelian free group. In these examples the holonomy representation takes values in $\mathrm{O}(2,1) \ltimes \mathbb{R}^3$ (this is always the case for $d = 3$ when $\pi_1(M)$ is not virtually polycyclic, see Fried and Goldman [1983]), hence M inherits a flat Lorentzian structure. Such manifolds are called *Margulis spacetimes*. They have a rich geometry and have been much studied since the 1990s, most prominently by Charette, Drumm, Goldman, Labourie, and Margulis. In particular, the questions of the topological tameness of Margulis spacetimes and of the existence of nice fundamental domains in $X = \mathbb{R}^3$ (bounded by piecewise linear objects called *crooked planes*) have received much attention: see e.g. Drumm [1992], Drumm and Goldman [1999], Charette, Drumm, and Goldman [2016], Choi and Goldman [2017], Danciger, Guéritaud, and Kassel [2016a,b]. See also Goldman, Labourie, and Margulis [2009], Abels, Margulis, and Soifer [2012], and Smilga [2016] for higher-dimensional analogues M with $\pi_1(M)$ a free group.

Following Danciger, Guéritaud, and Kassel [2016a,b] (see also Schlenker [2016]), a convenient point of view for understanding Margulis spacetimes is to regard them as "infinitesimal analogues" of complete AdS manifolds. In order to describe this point of view, we first briefly discuss the AdS case.

7.2 Complete AdS manifolds. As in Section 3.3, let $(G, X) = (\mathrm{PO}(2,2), \mathrm{AdS}^3)$, and view X as the group $\underline{G} = \mathrm{PSL}(2,\mathbb{R})$ and the identity component G_0 of G as $\underline{G} \times \underline{G}$ acting on $X \simeq \underline{G}$ by right and left multiplication. We consider (G, X)-manifolds of the form $M = \rho(\Gamma)\backslash X$ where $\Gamma \simeq \pi_1(M)$ is an infinite discrete group and $\rho = (\rho_L, \rho_R) : \Gamma \to \underline{G} \times \underline{G} \subset G$ a faithful representation through which Γ acts properly discontinuously and freely on X. Not all faithful and discrete $\rho = (\rho_L, \rho_R)$ yield properly discontinuous actions on X: e.g. if $\rho_L = \rho_R$, then ρ has a global fixed point, precluding properness. However, the following properness criteria hold. We denote by $\lambda(g) := \inf_{x \in \mathbb{H}^2} d_{\mathbb{H}^2}(x, g \cdot x) \geq 0$ the translation length of $g \in \underline{G}$ in $\mathbb{H}^2$.

Theorem 7.1 (Kassel [2009], Guéritaud, Guichard, Kassel, and Wienhard [2017a]). *Let $G = \mathrm{PO}(2,2)$ and $\underline{G} = \mathrm{PSL}(2,\mathbb{R})$. Consider a discrete group Γ and a representation*

$\rho = (\rho_L, \rho_R) : \Gamma \to \underline{G} \times \underline{G} \subset G$ with ρ_L convex cocompact. The following are equivalent, up to switching ρ_L and ρ_R in both (2) and (3):

(1) the action of Γ on $X = \mathrm{AdS}^3 \simeq \underline{G}$ via ρ is properly discontinuous;

(2) there exists $C < 1$ such that $\lambda(\rho_R(\gamma)) \le C\lambda(\rho_L(\gamma))$ for all $\gamma \in \Gamma$;

(3) there is a (ρ_L, ρ_R)-equivariant Lipschitz map $f : \mathbb{H}^2 \to \mathbb{H}^2$ with $\mathrm{Lip}(f) < 1$;

(4) Γ is Gromov hyperbolic and $\rho : \Gamma \to G \subset \mathrm{PGL}(4, \mathbb{R})$ is P_2-Anosov.

The equivalences (1) ⇔ (2) ⇔ (3), proved in Kassel [2009], have been generalized in Guéritaud and Kassel [2017] to $\underline{G} = \mathrm{PO}(n, 1)$ for any $n \ge 2$, allowing ρ_L to be geometrically finite instead of convex cocompact. These equivalences state that $\rho = (\rho_L, \rho_R)$ acts properly discontinuously on $X = \mathrm{AdS}^3 \simeq \underline{G}$ if and only if, up to switching the two factors, ρ_L is faithful and discrete and ρ_R is "uniformly contracting" with respect to ρ_L. The equivariant map f in (3) provides an explicit fibration in circles of $\rho(\Gamma)\backslash X$ over the hyperbolic surface $\rho_L(\Gamma)\backslash\mathbb{H}^2$, see Guéritaud and Kassel [ibid.]. We refer to Salein [2000], Guéritaud and Kassel [2017], Guéritaud, Kassel, and Wolff [2015], Deroin and Tholozan [2016], Danciger, Guéritaud, and Kassel [2018d], Lakeland and Leininger [2017] for many examples, to Tholozan [2017] for a classification in the compact AdS case, and to Kassel [2009] and Guéritaud and Kassel [2017] for links with the asymmetric metric on Teichmüller space introduced by Thurston [1986].

The equivalences (1) ⇔ (2) ⇔ (4), proved in Guéritaud, Guichard, Kassel, and Wienhard [2017a], generalize to $\underline{G} = \mathrm{PO}(n, 1)$, $\mathrm{PU}(n, 1)$, or $\mathrm{Sp}(n, 1)$; the Anosov condition is then expressed in $\mathrm{PGL}(2n+2, \mathbb{K})$ where $\mathbb{K}$ is $\mathbb{R}$, $\mathbb{C}$, or the quaternions. As an application, the set of holonomy representations of *complete* $(\underline{G} \times \underline{G}, \underline{G})$-structures on a compact manifold M is open in the set of holonomy representations of all possible $(\underline{G} \times \underline{G}, \underline{G})$-structures on M. By Tholozan [2015], it is also closed, which gives evidence for an open conjecture stating that all $(\underline{G} \times \underline{G}, \underline{G})$-structures on M should be complete (i.e. obtained as quotients of $\widetilde{\underline{G}}$).

7.3 Complete affine manifolds. We now go back to $(G, X) = (\mathrm{Aff}(\mathbb{R}^d), \mathbb{R}^d)$, looking for characterizations of holonomy representations of complete affine manifolds, i.e. representations to G yielding properly discontinuous actions on X.

We first note that any representation from a group Γ to the nonreductive Lie group $G = \mathrm{GL}(d, \mathbb{R}) \ltimes \mathbb{R}^d$ is of the form $\rho = (\rho_L, u)$ where $\rho_L : \Gamma \to \mathrm{GL}(d, \mathbb{R})$ (linear part) is a representation to $\mathrm{GL}(d, \mathbb{R})$ and $u : \Gamma \to \mathbb{R}^d$ (translational part) a ρ_L-cocycle, meaning $u(\gamma_1\gamma_2) = u(\gamma_1) + \rho_L(\gamma_1) \cdot u(\gamma_2)$ for all $\gamma_1, \gamma_2 \in \Gamma$.

We focus on the case $d = 3$ and ρ_L with values in $\mathrm{O}(2, 1)$. Let us briefly indicate how, following Danciger, Guéritaud, and Kassel [2016a,b], the Margulis spacetimes of

Section 7.1 are "infinitesimal versions" of the complete AdS manifolds of Section 7.2. Let $\underline{G} = \mathrm{O}(2,1)_0 \simeq \mathrm{PSL}(2,\mathbb{R})$ be the group of orientation-preserving isometries of $\mathbb{H}^2$. Its Lie algebra $\underline{\mathfrak{g}} \simeq \mathbb{R}^3$ is the set of "infinitesimal isometries" of $\mathbb{H}^2$, i.e. Killing vector fields on $\mathbb{H}^2$. Here are some properness criteria.

Theorem 7.2 (Goldman, Labourie, and Margulis [2009], Danciger, Guéritaud, and Kassel [2016a]). *Let $G = \mathrm{Aff}(\mathbb{R}^3)$ and $\underline{G} = \mathrm{O}(2,1)_0 \simeq \mathrm{PSL}(2,\mathbb{R})$. Consider a discrete group Γ and a representation $\rho = (\rho_L, u) : \Gamma \to \underline{G} \ltimes \underline{\mathfrak{g}} \subset G$ with ρ_L convex cocompact. The following are equivalent, up to replacing u by $-u$ in both (2) and (3):*

(1) the action of Γ on $X = \mathbb{R}^3 \simeq \underline{\mathfrak{g}}$ via $\rho = (\rho_L, u)$ is properly discontinuous;

(2) there exists $c < 0$ such that $\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\, \lambda(e^{u(\gamma)}\rho_L(\gamma)) \leq c\, \lambda(\rho_L(\gamma))$ for all $\gamma \in \Gamma$;

(3) there is a (ρ_L, u)-equivariant vector field Y on $\mathbb{H}^2$ with "lipschitz" constant < 0.

The equivalence (1) ⇔ (2) is a reinterpretation, based on Goldman and Margulis [2000], of a celebrated result of Goldman, Labourie, and Margulis [2009]. The equivalence (1) ⇔ (3) is proved in Danciger, Guéritaud, and Kassel [2016a].

These equivalences are "infinitesimal versions" of the equivalences (1) ⇔ (2) ⇔ (3) of Theorem 7.1. Indeed, as explained in Danciger, Guéritaud, and Kassel [ibid.], we can see the ρ_L-cocycle $u : \Gamma \to \underline{\mathfrak{g}}$ as an "infinitesimal deformation" of the holonomy representation ρ_L of the hyperbolic surface (or orbifold) $S = \rho_L(\Gamma)\backslash\mathbb{H}^2$; Condition (2) states that closed geodesics on S get uniformly shorter under this infinitesimal deformation. We can see a (ρ_L, u)-equivariant vector field Y on $\mathbb{H}^2$ as an "infinitesimal deformation" of the developing map of the hyperbolic surface S; Condition (3), which involves an unconventional notion of "lipschitz" constant, states that any two points of $\mathbb{H}^2$ get uniformly closer compared to their initial distance. Thus Theorem 7.2 states that $\rho = (\rho_L, u)$ acts properly discontinuously on $X = \mathbb{R}^3 \simeq \underline{\mathfrak{g}}$ if and only if the infinitesimal deformation u, up to replacing it by $-u$, is "uniformly contracting".

The vector field Y in (3) provides an explicit fibration in lines of the Margulis spacetime $\rho(\Gamma)\backslash X$ over the hyperbolic surface S, and this can be used to define a *geometric transition* from complete AdS manifolds to Margulis spacetimes, see Danciger, Guéritaud, and Kassel [ibid.].

In Theorem 7.1, the "uniform contraction" characterizing properness was in fact an Anosov condition, encoding strong dynamics on a certain flag variety. It is natural to expect that something similar should hold in the setting of Theorem 7.2. For this, a notion of *affine Anosov representation* to $\mathrm{O}(2,1) \ltimes \mathbb{R}^3$ was recently introduced by Ghosh [2017] and extended to $\mathrm{O}(n+1,n) \ltimes \mathbb{R}^d \subset \mathrm{Aff}(\mathbb{R}^d) = G$ for any $d = 2n+1 \geq 3$ by Ghosh and Treib [2017]; the definition is somewhat analogous to Section 4.1 but uses affine bundles and their sections. By Ghosh [2017] and Ghosh and Treib [2017], given a P_n-Anosov representation $\rho_L : \Gamma \to \mathrm{O}(n+1,n)$ and a ρ_L-cocycle $u : \Gamma \to \mathbb{R}^d$, the action of Γ on $X = \mathbb{R}^d$ via $\rho = (\rho_L, u)$ is properly discontinuous if and only if ρ is affine Anosov.

Theorem 7.2 was recently generalized in Danciger, Guéritaud, and Kassel [2018d] as follows: for $\underline{G} = \mathrm{O}(p,q)$ with $p,q \geq 1$, consider a discrete group Γ, a faithful and discrete representation $\rho_L : \Gamma \to \underline{G}$, and a ρ_L-cocycle $u : \Gamma \to \underline{\mathfrak{g}}$; then the action of Γ on $\underline{\mathfrak{g}}$ via $\rho = (\rho_L, u) : \Gamma \to \mathrm{Aff}(\underline{\mathfrak{g}})$ is properly discontinuous as soon as u satisfies a uniform contraction property in the pseudo-Riemannian hyperbolic space $\mathbb{H}^{p,q-1}$ of Section 6.2. This allowed for the construction of the first examples of irreducible complete affine manifolds M such that $\pi_1(M)$ is neither virtually polycyclic nor virtually free: $\pi_1(M)$ can in fact be any irreducible right-angled Coxeter group. It would be interesting to understand the links with a notion of affine Anosov representation in this setting.

8 Concluding remarks

By investigating the links between the geometry of (G,X)-structures on manifolds and the dynamics of their holonomy representations, we have discussed only a small part of a very active area of research.

We have described partial answers to Problem A for several types of model geometries (G,X). However, Problem A is still wide open in many contexts. As an illustration, let us mention two major open conjectures on closed affine manifolds (in addition to the Auslander conjecture of Section 7.1): the Chern conjecture states that if a closed d-manifold M admits an $(\mathrm{Aff}(\mathbb{R}^d), \mathbb{R}^d)$-structure, then its Euler characteristic must be zero; the Markus conjecture states that an $(\mathrm{Aff}(\mathbb{R}^d), \mathbb{R}^d)$-structure on M is complete if and only if its holonomy representation takes values in $\mathrm{SL}(d,\mathbb{R}) \ltimes \mathbb{R}^d$. See Klingler [2017] and references therein for recent progress on this.

We have seen that Anosov representations from Gromov hyperbolic groups to semisimple Lie groups provide a large class of representations answering Problem B. However, not much is known beyond them. One further class, generalizing Anosov representations to finitely generated groups Γ which are not necessarily Gromov hyperbolic, is the class of $\mathbb{P}(\mathbb{R}^d)$-convex cocompact representations to $\mathrm{PGL}(d,\mathbb{R})$ of Section 6.4; it would be interesting to understand this class better in the framework of Problem B, see Section 6.4 and Danciger, Guéritaud, and Kassel [2017, Appendix]. As another generalization of Anosov representations, it is natural to look for a class of representations of relatively hyperbolic groups to higher-rank semisimple Lie groups which would bear similarities to geometrically finite representations to rank-one groups, with cusps allowed: see Kapovich and Leeb [2017, § 5] for a conjectural picture. Partial work in this direction has been done in the convex projective setting, see Crampon and Marquis [2014].

To conclude, here are two open questions which we find particularly interesting.

Structural stability. Sullivan [1985] proved that a structurally stable, nonrigid subgroup of $G = \mathrm{PSL}(2, \mathbb{C})$ is always Gromov hyperbolic and convex cocompact in G. It is natural to ask if this may be extended to subgroups of higher-rank semisimple Lie groups G such as $\mathrm{PGL}(d, \mathbb{R})$ for $d \geq 3$, for instance with "convex cocompact" replaced by "Anosov". In Section 6.4 we saw that there exist nonrigid, structurally stable subgroups of $G = \mathrm{PGL}(d, \mathbb{R})$ which are *not* Gromov hyperbolic, namely groups that are $\mathbb{P}(\mathbb{R}^d)$-convex cocompact but not strongly $\mathbb{P}(\mathbb{R}^d)$-convex cocompact (Definitions 6.3 and 6.5). However, does a Gromov hyperbolic, nonrigid, structurally stable, discrete subgroup of G always satisfy some Anosov property?

Abstract groups admitting Anosov representations. Which linear hyperbolic groups admit Anosov representations to some semisimple Lie group? Classical examples include surface groups, free groups, and more generally rank-one convex cocompact groups, see Section 4.2. By Danciger, Guéritaud, and Kassel [2018a], all Gromov hyperbolic right-angled Coxeter groups (and all groups commensurable to them) admit Anosov representations, see Section 6.2. On the other hand, if a hyperbolic group admits an Anosov representation, then its Gromov flow (see Section 4.1) must satisfy strong dynamical properties, which may provide an obstruction: see Bridgeman, Canary, Labourie, and Sambarino [2015, end of § 1]. It would be interesting to have further concrete examples of groups admitting or not admitting Anosov representations.

References

Herbert Abels (2001). "Properly discontinuous groups of affine transformations: a survey". *Geom. Dedicata* 87.1-3, pp. 309–333. MR: 1866854 (cit. on p. 1161).

Herbert Abels, Gregory A. Margulis, and Gregory A. Soifer (2012). "The Auslander conjecture for dimension less than 7". arXiv: 1211.2525 (cit. on p. 1161).

Daniele Alessandrini and Qiongling Li (2018). "Projective structures with (quasi)-Hitchin holonomy". Preprint (cit. on p. 1153).

Louis Auslander (1964). "The structure of complete locally affine manifolds". *Topology* 3.suppl. 1, pp. 131–139. MR: 0161255 (cit. on p. 1161).

Samuel A. Ballas, Jeffrey Danciger, and Gye-Seon Lee (2018). "Convex projective structures on non-hyperbolic three-manifolds". *Geom. Topol.* 22.3, pp. 1593–1646. MR: 3780442 (cit. on p. 1142).

Thierry Barbot (2010). "Three-dimensional Anosov flag manifolds". *Geom. Topol.* 14.1, pp. 153–191. MR: 2578303 (cit. on pp. 1147, 1151).

– (2015). "Deformations of Fuchsian AdS representations are quasi-Fuchsian". *J. Differential Geom.* 101.1, pp. 1–46. MR: 3356068 (cit. on pp. 1145, 1157).

Thierry Barbot, Francesco Bonsante, Jeffrey Danciger, William M. Goldman, François Guéritaud, Fanny Kassel, Kirill Krasnov, Jean-Marc Schlenker, and Abdelghani Zeghib (2012). "Some open questions on anti-de Sitter geometry". arXiv: 1205.6103 (cit. on p. 1145).

Thierry Barbot and Quentin Mérigot (2012). "Anosov AdS representations are quasi-Fuchsian". *Groups Geom. Dyn.* 6.3, pp. 441–483. MR: 2961282 (cit. on pp. 1145, 1156, 1157).

Oliver Baues (2014). "The deformations of flat affine structures on the two-torus". In: *Handbook of Teichmüller theory. Vol. IV*. Vol. 19. IRMA Lect. Math. Theor. Phys. European Mathematical Society, Zürich, pp. 461–539. MR: 3289710 (cit. on pp. 1136, 1140).

Yves Benoist (1996). "Actions propres sur les espaces homogènes réductifs". *Ann. of Math. (2)* 144.2, pp. 315–347. MR: 1418901 (cit. on pp. 1148, 1154).

– (2004). "Convexes divisibles. I". In: *Algebraic groups and arithmetic*. Tata Inst. Fund. Res., Mumbai, pp. 339–374. MR: 2094116 (cit. on pp. 1142, 1143, 1150, 1158, 1159).

– (2005). "Convexes divisibles. III". *Ann. Sci. Éc. Norm. Sup. (4)* 38.5, pp. 793–832. MR: 2195260 (cit. on p. 1143).

– (2006). "Convexes divisibles. IV. Structure du bord en dimension 3". *Invent. Math.* 164.2, pp. 249–278. MR: 2218481 (cit. on pp. 1142, 1143, 1160).

– (2008). "A survey on divisible convex sets". In: *Geometry, analysis and topology of discrete groups*. Vol. 6. Adv. Lect. Math. (ALM). Int. Press, Somerville, MA, pp. 1–18. MR: 2464391 (cit. on p. 1142).

Lipman Bers (1960). "Simultaneous uniformization". *Bull. Amer. Math. Soc.* 66, pp. 94–97. MR: 0111834 (cit. on p. 1141).

Jairo Bochi, Rafael Potrie, and Andrés Sambarino (2018). "Anosov representations and dominated splittings". To appear in J. Eur. Math. Soc. (cit. on p. 1150).

Francis Bonahon (2002). "Geometric structures on 3-manifolds". In: *Handbook of geometric topology*. North-Holland, Amsterdam, pp. 93–164. MR: 1886669 (cit. on p. 1136).

Francis Bonahon and Guillaume Dreyer (2014). "Parameterizing Hitchin components". *Duke Math. J.* 163.15, pp. 2935–2975. MR: 3285861 (cit. on p. 1148).

Francesco Bonsante, Kirill Krasnov, and Jean-Marc Schlenker (2011). "Multi-black holes and earthquakes on Riemann surfaces with boundaries". *Int. Math. Res. Not. IMRN* 2011.3, pp. 487–552. MR: 2764871 (cit. on p. 1145).

Francesco Bonsante and Jean-Marc Schlenker (2012). "Fixed points of compositions of earthquakes". *Duke Math. J.* 161.6, pp. 1011–1054. MR: 2913100 (cit. on p. 1145).

Brian H. Bowditch (1993). "Geometrical finiteness for hyperbolic groups". *J. Funct. Anal.* 113.2, pp. 245–317. MR: 1218098 (cit. on p. 1140).

– (1998). "Spaces of geometrically finite representations". *Ann. Acad. Sci. Fenn. Math.* 23.2, pp. 389–414. MR: 1642122 (cit. on p. 1140).

Harrison Bray (2017). "Geodesic flow of nonstrictly convex Hilbert geometries". arXiv: 1710.06938 (cit. on p. 1143).

Martin Bridgeman, Richard Canary, François Labourie, and Andrés Sambarino (2015). "The pressure metric for Anosov representations". *Geom. Funct. Anal.* 25.4, pp. 1089–1179. MR: 3385630 (cit. on pp. 1146, 1148, 1165).

Jean-Philippe Burelle and Nicolaus Treib (2018). "Schottky groups and maximal representations". To appear in Geom. Dedicata (cit. on p. 1148).

Marc Burger, Alessandra Iozzi, François Labourie, and Anna Wienhard (2005). "Maximal representations of surface groups: symplectic Anosov structures". *Pure Appl. Math. Q.* 1.3, Special Issue: In memory of Armand Borel. Part 2, pp. 543–590. MR: 2201327 (cit. on p. 1147).

Marc Burger, Alessandra Iozzi, and Anna Wienhard (2010a). "Maximal representations and Anosov structures". In preparation (cit. on p. 1147).

– (2010b). "Surface group representations with maximal Toledo invariant". *Ann. of Math. (2)* 172.1, pp. 517–566. MR: 2680425 (cit. on p. 1137).

– (2014). "Higher Teichmüller spaces: from $\mathrm{SL}(2,\mathbb{R})$ to other Lie groups". In: *Handbook of Teichmüller theory. Vol. IV*. Vol. 19. IRMA Lect. Math. Theor. Phys. European Mathematical Society, Zürich, pp. 539–618. MR: 3289711 (cit. on pp. 1137, 1148).

Marc Burger and Maria Beatrice Pozzetti (2017). "Maximal representations, non-Archimedean Siegel spaces, and buildings". *Geom. Topol.* 21.6, pp. 3539–3599. MR: 3692972 (cit. on p. 1149).

Richard Canary, Michelle Lee, Andrés Sambarino, and Matthew Stover (2017). "Projective Anosov Schottky groups and strongly amalgam Anosov representations". *Geom. Topol.* 21.1. Appendix to the paper "Amalgam Anosov representations" by Canary, Lee, and Stover, pp. 240–248. MR: 3608713 (cit. on p. 1148).

Christophe Champetier (1994). "Petite simplification dans les groupes hyperboliques". *Ann. Fac. Sci. Toulouse Math. (6)* 3.2, pp. 161–221. MR: 1283206 (cit. on p. 1146).

Virginie Charette, Todd A. Drumm, and William M. Goldman (2016). "Proper affine deformations of the one-holed torus". *Transform. Groups* 21.4, pp. 953–1002. MR: 3569564 (cit. on p. 1161).

Suhyoung Choi and William M. Goldman (2005). "The deformation spaces of convex $\mathbb{RP}^2$-structures on 2-orbifolds". *Amer. J. Math.* 127.5, pp. 1019–1102. MR: 2170138 (cit. on pp. 1143, 1150, 1152).

– (2017). "Topological tameness of Margulis spacetimes". *Amer. J. Math.* 139.2, pp. 297–345. MR: 3636632 (cit. on p. 1161).

Suhyoung Choi, Gye-Seon Lee, and Ludovic Marquis (2016a). "Convex projective generalized Dehn filling". arXiv: 1611.02505 (cit. on p. 1142).

Suhyoung Choi, Gye-Seon Lee, and Ludovic Marquis (2016b). *Deformations of convex real projective structures on manifolds and orbifolds*. To appear in *Handbook of group actions*, Adv. Lect. Math., International Press (cit. on p. 1142).

Brian Collier, Nicolas Tholozan, and Jérémy Toulisse (2017). "The geometry of maximal representations of surface groups into $\mathrm{SO}_0(2, n)$". arXiv: 1702.08799 (cit. on p. 1153).

Daryl Cooper, Darren Long, and Stephan Tillmann (2018). "Deforming convex projective manifolds". *Geom. Topol.* 22.3, pp. 1349–1404. MR: 3780436 (cit. on p. 1160).

Mickaël Crampon (2014). "The geodesic flow of Finsler and Hilbert geometries". In: *Handbook of Hilbert geometry*. Vol. 22. IRMA Lect. Math. Theor. Phys. European Mathematical Society, Zürich, pp. 161–206. MR: 3329881 (cit. on p. 1143).

Mickaël Crampon and Ludovic Marquis (2014). "Finitude géométrique en géométrie de Hilbert". *Ann. Inst. Fourier (Grenoble)* 64.6, pp. 2299–2377. MR: 3331168 (cit. on pp. 1158, 1164).

Jeffrey Danciger, François Guéritaud, and Fanny Kassel (2016a). "Geometry and topology of complete Lorentz spacetimes of constant curvature". *Ann. Sci. Éc. Norm. Supér. (4)* 49.1, pp. 1–56. MR: 3465975 (cit. on pp. 1161–1163).

– (2016b). "Margulis spacetimes via the arc complex". *Invent. Math.* 204.1, pp. 133–193. MR: 3480555 (cit. on pp. 1161, 1162).

– (2017). "Convex cocompact actions in real projective geometry". arXiv: 1704.08711 (cit. on pp. 1152, 1153, 1156–1160, 1164).

– (2018a). "Convex cocompactness in pseudo-Riemannian hyperbolic spaces". *Geom. Dedicata* 192.1, Special issue: Geometries: A celebration of Bill Goldman's 60th birthday, pp. 87–126. MR: 3749424 (cit. on pp. 1153, 1156–1158, 1165).

– (2018b). "Examples and non-examples of convex cocompact actions in real projective space". Preprint (cit. on pp. 1159, 1160).

– (2018c). "Projective convex cocompactness for right-angled Coxeter groups". Preprint (cit. on pp. 1158, 1160).

– (2018d). "Proper affine actions for right-angled Coxeter groups". arXiv: 1804.03132 (cit. on pp. 1162, 1164).

Pallavi Dani (2017). "The large-scale geometry of right-angled Coxeter groups". Preprint (cit. on p. 1157).

Bertrand Deroin and Nicolas Tholozan (2016). "Dominating surface group representations by Fuchsian ones". *Int. Math. Res. Not. IMRN* 2016.13, pp. 4145–4166. MR: 3544632 (cit. on p. 1162).

Todd A. Drumm (1992). "Fundamental polyhedra for Margulis space-times". *Topology* 31.4, pp. 677–683. MR: 1191372 (cit. on p. 1161).

Todd A. Drumm and William M. Goldman (1999). "The geometry of crooked planes". *Topology* 38.2, pp. 323–351. MR: 1660333 (cit. on p. 1161).

Charles Ehresmann (1937). "Variétés localement projectives". *Enseign. Math.* 35, pp. 317–333 (cit. on p. 1135).

Vladimir Fock and Alexander Goncharov (2006). "Moduli spaces of local systems and higher Teichmüller theory". *Publ. Math. Inst. Hautes Études Sci.* 103, pp. 1–211. MR: 2233852 (cit. on pp. 1137, 1148).

Charles Frances (2005). "Lorentzian Kleinian groups". *Comment. Math. Helv.* 80.4, pp. 883–910. MR: 2182704 (cit. on p. 1152).

David Fried and William M. Goldman (1983). "Three-dimensional affine crystallographic groups". *Adv. in Math.* 47.1, pp. 1–49. MR: 689763 (cit. on p. 1161).

Sourav Ghosh (2017). "Anosov structures on Margulis spacetimes". *Groups Geom. Dyn.* 11.2, pp. 739–775. MR: 3668058 (cit. on p. 1163).

Sourav Ghosh and Nicolaus Treib (2017). "Affine Anosov representations and Proper actions". arXiv: 1711.09712 (cit. on p. 1163).

William M. Goldman (1990). "Convex real projective structures on compact surfaces". *J. Differential Geom.* 31.3, pp. 791–845. MR: 1053346 (cit. on pp. 1143, 1148, 1150, 1152).

– (2010). "Locally homogeneous geometric manifolds". In: *Proceedings of the International Congress of Mathematicians 2010. Vol. II.* World Scientific, pp. 717–744. MR: 2827816 (cit. on pp. 1135, 1139).

– (2018a). *Flat affine, projective and conformal structures on manifolds: a historical perspective*. To appear in Geometry in history, Springer-Verlag (cit. on p. 1135).

– (2018b). *Geometric structures on manifolds*. In preparation for the American Mathematical Society, Graduate Studies in Mathematics (cit. on pp. 1134, 1135, 1139).

William M. Goldman, François Labourie, and Gregory A. Margulis (2009). "Proper affine actions and geodesic flows of hyperbolic surfaces". *Ann. of Math. (2)* 170.3, pp. 1051–1083. MR: 2600870 (cit. on pp. 1161, 1163).

William M. Goldman and Gregory A. Margulis (2000). "Flat Lorentz 3-manifolds and cocompact Fuchsian groups". In: *Crystallographic groups and their generalizations (Kortrijk, 1999)*. Vol. 262. Contemp. Math. American Mathematical Society, Providence, RI, pp. 135–145. MR: 1796129 (cit. on p. 1163).

Mikhail Gromov (1987). "Hyperbolic groups". In: *Essays in group theory*. Vol. 8. Math. Sci. Res. Inst. Publ. Springer, New York, pp. 75–263. MR: 919829 (cit. on p. 1146).

François Guéritaud, Olivier Guichard, Fanny Kassel, and Anna Wienhard (2017a). "Anosov representations and proper actions". *Geom. Topol.* 21.1, pp. 485–584. MR: 3608719 (cit. on pp. 1149, 1150, 1154, 1161, 1162).

– (2017b). "Compactification of certain Clifford-Klein forms of reductive homogeneous spaces". *Michigan Math. J.* 66.1, pp. 49–84. MR: 3619735 (cit. on p. 1154).

François Guéritaud and Fanny Kassel (2017). "Maximally stretched laminations on geometrically finite hyperbolic manifolds". *Geom. Topol.* 21.2, pp. 693–840. MR: 3626591 (cit. on p. 1162).

François Guéritaud, Fanny Kassel, and Maxime Wolff (2015). "Compact anti-de Sitter 3-manifolds and folded hyperbolic structures on surfaces". *Pacific J. Math.* 275.2, pp. 325–359. MR: 3347373 (cit. on p. 1162).

Olivier Guichard (2017). "Groupes convexes-cocompacts en rang supérieur [d'après Labourie, Kapovich, Leeb, Porti, ...]" In: Séminaire Bourbaki (cit. on pp. 1147, 1155).

Olivier Guichard, Fanny Kassel, and Anna Wienhard (2015). "Tameness of Riemannian locally symmetric spaces arising from Anosov representations". arXiv: 1508.04759 (cit. on p. 1154).

Olivier Guichard and Anna Wienhard (2008). "Convex foliated projective structures and the Hitchin component for $\mathrm{PSL}_4(\mathbf{R})$". *Duke Math. J.* 144.3, pp. 381–445. MR: 2444302 (cit. on pp. 1152, 1153).

– (2012). "Anosov representations: domains of discontinuity and applications". *Invent. Math.* 190.2, pp. 357–438. MR: 2981818 (cit. on pp. 1145–1149, 1151–1154, 1158).

– (2016). "Positivity and higher Teichmüller theory". In: Proceedings of the European Congress of Mathematics 2016 (cit. on p. 1149).

Nigel J. Hitchin (1992). "Lie groups and Teichmüller space". *Topology* 31.3, pp. 449–473. MR: 1174252 (cit. on pp. 1147, 1148).

Michael Kapovich and Bernhard Leeb (2015). "Finsler bordifications of symmetric and certain locally symmetric spaces". arXiv: 1505.03593 (cit. on p. 1154).

– (2017). "Discrete isometry groups of symmetric spaces". arXiv: 1703.02160 (cit. on pp. 1147, 1149, 1155, 1164).

Michael Kapovich, Bernhard Leeb, and Joan Porti (2014a). "A Morse Lemma for quasi-geodesics in symmetric spaces and euclidean buildings". arXiv: 1411.4176 (cit. on pp. 1148–1150).

– (2014b). "Morse actions of discrete groups on symmetric space". arXiv: 1403.7671 (cit. on p. 1149).

– (2016). "Some recent results on Anosov representations". *Transform. Groups* 21.4, pp. 1105–1121. MR: 3569569 (cit. on pp. 1147, 1155).

– (2017). "Anosov representations: dynamical and geometric characterizations". *Eur. J. Math.* 3.4, pp. 808–898. MR: 3736790 (cit. on pp. 1147, 1155).

– (2018). "Dynamics on flag manifolds: domains of proper discontinuity and cocompactness". *Geom. Topol.* 22.1, pp. 157–234. MR: 3720343 (cit. on pp. 1152, 1154).

Fanny Kassel (2009). "Quotients compacts d'espaces homogènes réels ou *p*-adiques". PhD thesis. Université Paris-Sud XI (cit. on pp. 1154, 1161, 1162).

Fanny Kassel and Rafael Potrie (2018). "Eigenvalue gaps for hyperbolic groups and semigroups". Preprint (cit. on p. 1150).

Bruce Kleiner and Bernhard Leeb (2006). "Rigidity of invariant convex sets in symmetric spaces". *Invent. Math.* 163.3, pp. 657–676. MR: 2207236 (cit. on p. 1155).

Bruno Klingler (2017). "Chern's conjecture for special affine manifolds". *Ann. of Math. (2)* 186.1, pp. 69–95. MR: 3665000 (cit. on p. 1164).

Toshiyuki Kobayashi (1996). "Criterion for proper actions on homogeneous spaces of reductive groups". *J. Lie Theory* 6.2, pp. 147–163. MR: 1424629 (cit. on p. 1154).

Toshiyuki Kobayashi and Taro Yoshino (2005). "Compact Clifford-Klein forms of symmetric spaces—revisited". *Pure Appl. Math. Q.* 1.3, Special Issue: In memory of Armand Borel. Part 2, pp. 591–663. MR: 2201328 (cit. on pp. 1135, 1154).

Jean-Louis Koszul (1968). "Déformations de connexions localement plates". *Ann. Inst. Fourier (Grenoble)* 18.fasc. 1, pp. 103–114. MR: 0239529 (cit. on p. 1143).

François Labourie (2006). "Anosov flows, surface groups and curves in projective space". *Invent. Math.* 165.1, pp. 51–114. MR: 2221137 (cit. on pp. 1137, 1145, 1147, 1149).

François Labourie (2008). "Cross ratios, Anosov representations and the energy functional on Teichmüller space". *Ann. Sci. Éc. Norm. Sup. (4)* 41.3, pp. 439–471. MR: 2482204 (cit. on p. 1148).

Grant S. Lakeland and Christopher J. Leininger (2017). "Strict contractions and exotic $SO_0(d,1)$ quotients". *J. Lond. Math. Soc.* 96.3, pp. 642–662. MR: 3742437 (cit. on p. 1162).

Gye-Seon Lee and Ludovic Marquis (2018). "Anti-de Sitter strictly GHC-regular groups which are not lattices". To appear in Trans. Amer. Math. Soc (cit. on p. 1156).

Gye-Seon Lee and Tengren Zhang (2017). "Collar lemma for Hitchin representations". *Geom. Topol.* 21.4, pp. 2243–2280. MR: 3654108 (cit. on p. 1149).

Gregory A. Margulis (1984). "Complete affine locally flat manifolds with a free fundamental group". *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* 134. Automorphic functions and number theory, II, pp. 190–205. MR: 741860 (cit. on p. 1161).

Frédéric Mathéus (1991). "Flot géodésique et groupes hyperboliques d'après M. Gromov". In: *Séminaire de Théorie Spectrale et Géométrie, No. 9, Année 1990–1991*. Vol. 9. Sémin. Théor. Spectr. Géom. Univ. Grenoble I, Saint-Martin-d'Hères, pp. 67–87. MR: 1715930 (cit. on p. 1146).

Geoffrey Mess (2007). "Lorentz spacetimes of constant curvature". *Geom. Dedicata* 126. Preprint from 1990, pp. 3–45. MR: 2328921 (cit. on pp. 1144, 1145, 1151).

John Milnor (1977). "On fundamental groups of complete affinely flat manifolds". *Advances in Math.* 25.2, pp. 178–187. MR: 0454886 (cit. on p. 1161).

Igor Mineyev (2005). "Flows and joins of metric spaces". *Geom. Topol.* 9, pp. 403–482. MR: 2140987 (cit. on p. 1146).

Mark Pollicott and Richard Sharp (2017). "Higher Teichmüller theory for surface groups and shifts of finite type". Preprint (cit. on p. 1148).

Jean-François Quint (2005). "Groupes convexes cocompacts en rang supérieur". *Geom. Dedicata* 113, pp. 1–19. MR: 2171296 (cit. on p. 1155).

François Salein (2000). "Variétés anti-de Sitter de dimension 3 exotiques". *Ann. Inst. Fourier (Grenoble)* 50.1, pp. 257–284. MR: 1762345 (cit. on p. 1162).

Jean-Marc Schlenker (2016). "Variétés lorentziennes plates vues comme limites de variétés anti-de Sitter [d'après Danciger, Guéritaud et Kassel]". *Astérisque* 380, Séminaire Bourbaki. Vol. 2014/2015, Exp. No. 1103, 475–497. MR: 3522183 (cit. on p. 1161).

Ilia Smilga (2016). "Proper affine actions: a sufficient criterion". arXiv: 1612.08942 (cit. on p. 1161).

Tobias Strubel (2015). "Fenchel–Nielsen coordinates for maximal representations". *Geom. Dedicata* 176, pp. 45–86. MR: 3347573 (cit. on p. 1148).

Dennis Sullivan (1979). "The density at infinity of a discrete group of hyperbolic motions". *Publ. Math. Inst. Hautes Études Sci.* 50, pp. 171–202. MR: 556586 (cit. on p. 1141).

– (1985). "Quasiconformal homeomorphisms and dynamics. II. Structural stability implies hyperbolicity for Kleinian groups". *Acta Math.* 155.3-4, pp. 243–260. MR: 806415 (cit. on pp. 1141, 1165).

Nicolas Tholozan (2015). "Sur la complétude de certaines variétés pseudo-riemanniennes localement symétriques". *Ann. Inst. Fourier (Grenoble)* 65.5, pp. 1921–1952. MR: 3449201 (cit. on p. 1162).

– (2017). "Dominating surface group representations and deforming closed anti-de Sitter 3-manifolds". *Geom. Topol.* 21.1, pp. 193–214. MR: 3608712 (cit. on p. 1162).

William P. Thurston (1980). *The geometry and topology of three-manifolds*. Lecture notes (cit. on pp. 1135, 1140).

– (1986). "Minimal stretch maps between hyperbolic surfaces". arXiv: math/9801039 (cit. on p. 1162).

Ernest B. Vinberg (1971). "Discrete linear groups generated by reflections". *Mathematics of the USSR-Izvestiya* 5.5, pp. 1083–1119 (cit. on p. 1158).

Anna Wienhard (2006). "The action of the mapping class group on maximal representations". *Geom. Dedicata* 120.1, pp. 179–191. MR: 2252900 (cit. on p. 1148).

– (2018). "An invitation to higher Teichmüller theory". In: *Proceedings of the International Congress of Mathematicians 2018* (cit. on pp. 1137, 1148).

Edward Witten (1988). "$2+1$-dimensional gravity as an exactly soluble system". *Nuclear Phys. B* 311.1, pp. 46–78. MR: 974271 (cit. on p. 1143).

Tengren Zhang (2015). "Degeneration of Hitchin representations along internal sequences". *Geom. Funct. Anal.* 25.5, pp. 1588–1645. MR: 3426063 (cit. on p. 1148).

Andrew Zimmer (2017). "Projective Anosov representations, convex cocompact actions, and rigidity". arXiv: 1704.08582 (cit. on pp. 1152, 1153, 1158).

Received 2017-12-31.

Fanny Kassel
CNRS and Institut des Hautes Études Scientifiques
Laboratoire Alexander Grothendieck
35 route de Chartres
91440 Bures-sur-Yvette
France
kassel@ihes.fr

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1171–1192)

SPATIAL REFINEMENTS AND KHOVANOV HOMOLOGY

Robert Lipshitz and Sucharit Sarkar

Abstract

We review the construction and context of a stable homotopy refinement of Khovanov homology.

1 Introduction

While studying critical points and geodesics, Morse [1925, 1930, 1996] introduced what is now called *Morse theory*—using functions for which the second derivative test does not fail (*Morse functions*) to decompose manifolds into simpler pieces. The finite-dimensional case was further developed by many authors (see Bott [1980] for a survey of the history), and an infinite-dimensional analogue introduced by Palais and Smale [1964], Palais [1963], and Smale [1964]. In both cases, a Morse function f on M leads to a chain complex $C_*(f)$ generated by the critical points of f. This chain complex satisfies the *fundamental theorem of Morse homology*: its homology $H_*(f)$ is isomorphic to the singular homology of M. This is both a feature and a drawback: it means that one can use information about the topology of M to deduce the existence of critical points of f, but also implies that $C_*(f)$ does not see the smooth topology of M. (See Milnor [1963, 1965] for an elegant account of the subject's foundations and some of its applications.)

Much later, Floer [1988c,a,b] introduced some new examples of infinite-dimensional, Morse-like theories. Unlike Palais-Smale's Morse theory, in which the descending manifolds of critical points are finite-dimensional, in Floer's setting both ascending and descending manifolds are infinite-dimensional. Also unlike Palais-Smale's setting, Floer's homology groups are not isomorphic to singular homology of the ambient space (though the singular homology acts on them). Indeed, most Floer (co)homology theories seem to have no intrinsic cup product operation, and so are unlikely to be the homology of any natural space.

RL was supported by NSF DMS-1642067. SS was supported by NSF DMS-1643401.
MSC2010: primary 57M25; secondary 55P42.
Keywords: Framed flow categories, Cohen-Jones-Segal construction, Khovanov stable homotopy type, Burnside category.

Cohen, J. Jones, and Segal [1995] proposed that although Floer homology is not the homology of a space, it could be the homology of some associated spectrum (or pro-spectrum), and outlined a construction, under restrictive hypotheses, of such an object. While they suggest that these spectra might be determined by the ambient, infinite-dimensional manifold together with its polarization (a structure which seems ubiquitous in Floer theory), their construction builds a CW complex cell-by-cell, using the moduli spaces appearing in Floer theory. (We review their construction in Section 2.4. Steps towards describing Floer homology in terms of a polarized manifold have been taken by Lipyanskiy [n.d.].) Although the Cohen-Jones-Segal approach has been stymied by analytic difficulties, it has inspired other constructions of stable homotopy refinements of various Floer homologies and related invariants; see Furuta [2001], Bauer and Furuta [2004], Bauer [2004], Manolescu [2003], Kronheimer and Manolescu [n.d.], Douglas [n.d.], Cohen [2010, 2009], Kragh [n.d., 2013], Abouzaid and Kragh [2016], Khandhawit [2015a,b], Sasahira [n.d.], and Khandhawit, Lin, and Sasahira [n.d.].

From the beginning, Floer homologies have been used to define invariants of objects in low-dimensional topology—3-manifolds, knots, and so on. In a slightly different direction, Khovanov [2000] defined another knot invariant, which he calls $\mathfrak{sl}_2$ *homology* and everyone else calls *Khovanov homology*, whose graded Euler characteristic is the Jones polynomial from V. Jones [1985]. (See Bar-Natan [2002] for a friendly introduction.) While it looks formally similar to Floer-type invariants, Khovanov homology is defined combinatorially. No obvious infinite-dimensional manifold or functional is present. Still, Seidel and Smith [2006] (inspired by earlier work of Khovanov and Seidel [2002] and others) gave a conjectural reformulation of Khovanov homology via Floer homology. Over $\mathbb{Q}$, the isomorphism between Seidel-Smith's and Khovanov's invariants was recently proved by Abouzaid and Smith [n.d.]. Manolescu [2007] gave an extension of the reformulation to $\mathfrak{sl}_n$ homology constructed by Khovanov and Rozansky [2008].

Inspired by this history, Lipshitz and Sarkar [2014a,c,b] gave a combinatorial definition of a spectrum refining Khovanov homology, and studied some of its properties. This circle of ideas was further developed in Lipshitz, Ng, and Sarkar [2015] and in Lawson, Lipshitz, and Sarkar [n.d.(a),(b)], and extended in many directions by other authors (see Section 3.3). Another approach to a homotopy refinement was given by Everitt and Turner [2014], though it turns out their invariant is determined by Khovanov homology, cf. Everitt, Lipshitz, Sarkar, and Turner [2016]. Inspired by a different line of inquiry, Hu, D. Kriz, and I. Kriz [2016] also gave a construction of a Khovanov stable homotopy type. Lawson, Lipshitz, and Sarkar [n.d.(a)] show that the two constructions give homotopy equivalent spectra, perhaps suggesting some kind of uniqueness.

Most of this note is an outline of a construction of a Khovanov homotopy type, following Lawson, Lipshitz, and Sarkar [ibid.] and Hu, D. Kriz, and I. Kriz [2016], with an emphasis on the general question of stable homotopy refinements of chain complexes. In the last two sections, we briefly outline some of the structure and uses of the homotopy type (Section 3.3) and some questions and speculation (Section 3.4). Another exposition of some of this material can be found in Lawson, Lipshitz, and Sarkar [2017].

Acknowledgments. We are deeply grateful to our collaborator Tyler Lawson, whose contributions and perspective permeate the account below. We also thank Mohammed Abouzaid, Ciprian Manolescu, and John Pardon for comments on a draft of this article.

2 Spatial refinements

The spatial refinement problem can be summarized as follows.

Start with a chain complex C_* with a distinguished, finite basis, arising in some interesting setting. Incorporating more information from the setting, construct a based CW complex (or spectrum) whose reduced cellular chain complex, after a shift, is isomorphic to C_* with cells corresponding to the given basis.

A result of Carlsson [1981] implies that there is no universal solution to the spatial refinement problem, i.e., no functor S from chain complexes (supported in large gradings, say) to CW complexes so that the composition of S and the reduced cellular chain complex functor is the identity (cf. Prasma et al. [n.d.]). Specifically, for $G = \mathbb{Z}/2\times\mathbb{Z}/2$ he defines a module P over $\mathbb{Z}[G]$ so that there is no G-equivariant Moore space $M(P,n)$ for any n. If C_* is a free resolution of P over $\mathbb{Z}[G]$ then $S(C_*)$ would be such a Moore space, a contradiction.

Thus, spatially refining C_* requires context-specific work. This section gives general frameworks for such spatial refinements, and the next section has an interesting example of one.

2.1 Linear and cubical diagrams. Let C_* be a freely and finitely generated chain complex with a given basis. After shifting we may assume C_* is supported in gradings $0,\dots,n$. Let $[n+1]$ be the category with objects $0,1,\dots,n$ and a unique morphism $i\to j$ if $i\geq j$. Let $\mathcal{B}(\mathbb{Z})$ denote the category of finitely generated free abelian groups, with objects finite sets and $\operatorname{Hom}_{\mathcal{B}(\mathbb{Z})}(S,T)$ the set of linear maps $\mathbb{Z}\langle S\rangle\to\mathbb{Z}\langle T\rangle$ or, equivalently, $T\times S$ matrices of integers. Then C_* may be viewed as a functor F from $[n+1]$ to $\mathcal{B}(\mathbb{Z})$ subject to the condition that F sends any length two arrow (that is, a morphism $i\to j$ with $i-j=2$) to the zero map. Given such a functor $F\colon [n+1]\to\mathcal{B}(\mathbb{Z})$, that is, a *linear diagram*

$$\mathbb{Z}\langle F(n)\rangle\to\mathbb{Z}\langle F(n-1)\rangle\to\cdots\to\mathbb{Z}\langle F(1)\rangle\to\mathbb{Z}\langle F(0)\rangle \tag{2-1}$$

with every composition the zero map, we obtain a chain complex C_* by shifting the gradings, $C_i = \mathbb{Z}\langle F(i)\rangle[i]$, and letting $\partial_i = F(i \to i-1)$. This construction is functorial. That is, if $\mathscr{B}(\mathbb{Z})^{[n+1]}_{\bullet}$ denotes the full subcategory of the functor category $\mathscr{B}(\mathbb{Z})^{[n+1]}$ generated by those functors which send every length two arrow to the zero map, and if Kom denotes the category of chain complexes, then the above construction is a functor $\mathsf{ch}\colon \mathscr{B}(\mathbb{Z})^{[n+1]}_{\bullet} \to \mathsf{Kom}$. Indeed, it would be reasonable to call an element of $\mathscr{B}(\mathbb{Z})^{[n+1]}_{\bullet}$ a *chain complex in* $\mathscr{B}(\mathbb{Z})$.

A linear diagram $F \in \mathscr{B}(\mathbb{Z})^{[n+1]}_{\bullet}$ may also be viewed as a *cubical diagram* $G\colon [2]^n \to \mathscr{B}(\mathbb{Z})$ by setting

$$G(v) = \begin{cases} F(i) & \text{if } v = (\underbrace{0,\dots,0}_{n-i},\underbrace{1,\dots,1}_{i}) \\ \varnothing & \text{otherwise.} \end{cases} \tag{2-2}$$

On morphisms, G is either zero or induced from F, as appropriate. Conversely, a cubical diagram $G \in \mathscr{B}(\mathbb{Z})^{[2]^n}$ gives a linear diagram $F \in \mathscr{B}(\mathbb{Z})^{[n+1]}_{\bullet}$ by setting $F(i) = \coprod_{|v|=i} G(v)$, where $|v|$ denotes the number of 1's in v. The component of $F(i+1 \to i)$ from $\mathbb{Z}\langle G(u)\rangle \subset \mathbb{Z}\langle F(i+1)\rangle$ to $\mathbb{Z}\langle G(v)\rangle \subset \mathbb{Z}\langle F(i)\rangle$ is

$$\begin{cases} (-1)^{u_1+\dots+u_{k-1}} G(u \to v) & \text{if } u - v = \widehat{e}_k, \text{ the } k^{\text{th}} \text{ unit vector,} \\ 0 & \text{if } u - v \text{ is not a unit vector.} \end{cases} \tag{2-3}$$

These give functors $\mathscr{B}(\mathbb{Z})^{[n+1]}_{\bullet} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathscr{B}(\mathbb{Z})^{[2]^n}$ with $\beta \circ \alpha = \mathrm{Id}$.

The composition $\mathsf{ch} \circ \beta\colon \mathscr{B}(\mathbb{Z})^{[2]^n} \to \mathsf{Kom}$ is the *totalization* Tot, and may be viewed as an iterated mapping cone. Up to chain homotopy equivalence, one can also construct Tot using homotopy colimits. Define a category $[2]_+$ by adjoining a single object $*$ to $[2]$ and a single morphism $1 \to *$; let $[2]^n_+ = ([2]_+)^n$. Given $G \in \mathscr{B}(\mathbb{Z})^{[2]^n}$, by treating abelian groups as chain complexes supported in homological grading zero, we get an associated cubical diagram $A\colon [2]^n \to \mathsf{Kom}$. Extend A to a diagram $A_+\colon [2]^n_+ \to \mathsf{Kom}$ by setting

$$A_+(v) = \begin{cases} A(v) & \text{if } v \in [2]^n \\ 0 & \text{otherwise.} \end{cases} \tag{2-4}$$

Then the totalization of G is the *homotopy colimit* of A_+. (See Segal [1974], Bousfield and Kan [1972], and Vogt [1973].)

2.2 Spatial refinements of diagrams of abelian groups. As a next step, given a finitely generated chain complex represented by a functor $F\colon [n+1] \to \mathscr{B}(\mathbb{Z})$ we wish to construct a based cell complex with cells in dimensions $N, \dots, N+n$ whose reduced cellular complex—with distinguished basis given by the cells—is isomorphic to the given complex shifted up by N.

Let $\mathcal{T}(S^N)$ be the category with objects finite sets and morphisms $\mathrm{Hom}_{\mathcal{T}(S^N)}(S,T)$ the set of all based maps $\bigvee_S S^N \to \bigvee_T S^N$ between wedges of N-dimensional spheres; applying reduced N^{th} homology to the morphisms produces a functor, also denoted $\widetilde{H}_N$, from $\mathcal{T}(S^N)$ to $\mathscr{B}(\mathbb{Z})$. A *strict N-dimensional spatial lift* of F is a functor $P\colon [n+1] \to \mathcal{T}(S^N)$ satisfying $\widetilde{H}_N \circ P = F$ and P is the constant map on any length two arrow in $[n+1]$, i.e., a *strict chain complex in $\mathcal{T}(S^N)$ lifting F*. Just as morphisms in $\mathscr{B}(\mathbb{Z})$ are matrices, if we replace S^N by the sphere spectrum $\mathbb{S}$, we may view a morphism in $\mathcal{T}(\mathbb{S})$ as a matrix of maps $\mathbb{S} \to \mathbb{S}$ by viewing $\bigvee_S \mathbb{S}$ as a coproduct and $\bigvee_T \mathbb{S}$ as a product.

Given such a linear diagram P, we can construct a based cell complex by taking mapping cones and suspending sequentially, cf. Cohen, J. Jones, and Segal [1995, §5]. If CW denotes the category of based cell complexes, then P induces a diagram $X\colon [n+1] \to \mathsf{CW}$,

$$X(n) \xrightarrow{f_n} X(n-1) \xrightarrow{f_{n-1}} \cdots \xrightarrow{f_2} X(1) \xrightarrow{f_1} X(0) \tag{2-5}$$

with every composition the constant map. Since $f_1 \circ f_2$ is the constant map, there is an induced map $g_1\colon \Sigma X(2) \to \mathrm{Cone}(f_1)$ from the reduced suspension to the reduced cone. Then we get a diagram $Y\colon [n] \to \mathsf{CW}$,

$$Y(n-1) = \Sigma X(n) \xrightarrow{\Sigma f_n} \cdots \xrightarrow{\Sigma f_3} Y(1) = \Sigma X(2) \xrightarrow{g_1} Y(0) = \mathrm{Cone}(f_n). \tag{2-6}$$

Take the mapping cone of g_1 and suspend to get a diagram $Z\colon [n-1] \to \mathsf{CW}$ and so on. The reduced cellular chain complex of the final CW complex is the original chain complex, shifted up by N. This construction is also functorial: if $\mathcal{T}(S^N)^{[n+1]}_{\bullet}$ is the full subcategory generated by the functors which send every length two arrow to the constant map, then the construction is a functor $\mathcal{T}(S^N)^{[n+1]}_{\bullet} \to \mathsf{CW}$. The construction can also be carried out in a single step. Construct a category $[n+1]_+$ by adjoining a single object $*$ and a unique morphism $i \to *$ for all $i \neq 0$. Extend $X\colon [n+1] \to \mathsf{CW}$ to $X_+\colon [n+1]_+ \to \mathsf{CW}$ by sending $*$ to a point and take the homotopy colimit of X_+.

A linear diagram $P \in \mathcal{T}(S^N)^{[n+1]}_{\bullet}$ produces a cubical diagram $Q\colon [2]^n \to \mathcal{T}(S^N)$ by the analogue of Equation (2-2). There is a *totalization* functor $\mathsf{Tot}\colon \mathcal{T}(S^N)^{[2]^n} \to \mathsf{CW}$

extending the functor $\mathscr{T}(S^N)_\bullet^{[n+1]} \to \mathsf{CW}$ so that

$$\begin{array}{ccccc} \mathscr{T}(S^N)_\bullet^{[n+1]} & \longrightarrow & \mathscr{T}(S^N)^{[2]^n} & \longrightarrow & \mathsf{CW} \\ \downarrow & & \downarrow & & \downarrow \widetilde{C}_*^{\mathrm{cell}}[-N] \\ \mathscr{B}(\mathbb{Z})_\bullet^{[n+1]} & \longrightarrow & \mathscr{B}(\mathbb{Z})^{[2]^n} & \longrightarrow & \mathsf{Kom}. \end{array} \tag{2-7}$$

commutes. The totalization functor is defined as an iterated mapping cone or as a homotopy colimit of an extension of Q analogous to Equation (2-4).

2.3 Lax spatial refinements. Instead of working with strict functors as in the previous section, sometimes is it more convenient to work with lax functors. A *lax* or *homotopy coherent* or $(\infty, 1)$ *functor* $F: \mathscr{C} \to \mathsf{Top}$ is a diagram that commutes up to homotopies which are specified, and the homotopies themselves commute up to higher homotopies which are also specified, and so on; for details see Vogt [1973], Cordier [1982], and Lurie [2009a]. More precisely, F consists of based topological spaces $F(x)$ for $x \in \mathscr{C}$, and higher homotopy maps $F(f_n, \dots, f_1)\colon [0,1]^{n-1} \times F(x_0) \to F(x_n)$ for composable morphisms $x_0 \xrightarrow{f_1} \cdots \xrightarrow{f_n} x_n$ with certain boundary conditions and restrictions involving basepoints and identity morphisms; the case $n = 1$ is maps corresponding to the arrows in the diagram, $n = 2$ is homotopies corresponding to pairs of composable arrows, etc. A strict functor may be viewed as a lax functor. Let $\mathfrak{h}\mathsf{Top}^{\mathscr{C}}$ denote the category of lax functors $\mathscr{C} \to \mathsf{Top}$, with morphisms given by lax functors $\mathscr{C} \times [2] \to \mathsf{Top}$. (There are also higher morphisms corresponding to lax functors $\mathscr{C} \times [n] \to \mathsf{Top}$.)

There is a notion of a lax functor to $\mathscr{T}(S^N)$ induced from the notion of lax functors to Top. Let $\mathfrak{h}\mathscr{T}(S^N)_\bullet^{[n+1]}$ be the subcategory of $\mathfrak{h}\mathscr{T}(S^N)^{[n+1]}$ consisting of those objects (respectively, morphisms) F such that $F\big(x_0 \xrightarrow{f_1} \cdots \xrightarrow{f_n} x_n\big)$ is the constant map to the basepoint for any string of morphisms in $[n+1]$ (respectively, $[n+1] \times [2]$) with some f_i of length ≥ 2 in $[n+1]$ (respectively, $[n+1] \times \{0,1\}$). We call such functors *chain complexes in* $\mathscr{T}(S^N)$. (In Cohen-Jones-Segal's language, chain complexes in $\mathscr{T}(S^N)$ are functors $\mathscr{J}_0^n \to \mathscr{T}_*$.)

If one starts with a chain complex $F \in \mathscr{B}(\mathbb{Z})_\bullet^{[n+1]}$ and wishes to refine it to a based cell complex, instead of constructing a strict N-dimensional spatial lift in $\mathscr{T}(S^N)_\bullet^{[n+1]}$, it is enough to construct a *lax* N*-dimensional spatial lift*, that is, a functor $P \in \mathfrak{h}\mathscr{T}(S^N)_\bullet^{[n+1]}$ with $\widetilde{H}_N \circ P = F$. Such a P produces a cell complex by adjoining a basepoint to get a lax diagram $[n+1]_+ \to \mathsf{CW}$ and then taking homotopy colimits. Alternatively, we may convert P to a lax cubical diagram $Q \in \mathfrak{h}\mathscr{T}(S^N)^{[2]^n}$ and proceed as before. The iterated mapping cone construction becomes intricate since the associated diagram $X\colon [2]^n \to \mathsf{CW}$ is lax. So, extend to a lax diagram $X_+\colon [2]^n_+ \to \mathsf{CW}$ as before and then take its homotopy colimit. This generalization to the lax set-up remains functorial and the analogue of Diagram (2-7) still commutes.

2.4 Framed flow categories. Cohen, J. Jones, and Segal [1995] first proposed lax spatial refinements of diagrams $F\colon [n+1] \to \mathscr{B}(\mathbb{Z})$ via framed flow categories, using the Pontryagin-Thom construction. A *framed flow category* is an abstraction of the gradient flows of a Morse-Smale function. Concretely, a framed flow category $\mathscr{C}$ consists of:

1. A finite set of *objects* $\mathrm{Ob}(\mathscr{C})$ and a *grading* $\mathrm{gr}\colon \mathrm{Ob}(\mathscr{C}) \to \mathbb{Z}$. After translating, we may assume the gradings lie in $[0,n]$.
2. For $x, y \in \mathscr{C}$ with $\mathrm{gr}(x) - \mathrm{gr}(y) - 1 = k$, a *morphism set* $\mathfrak{M}(x,y)$ which is a k-dimensional $\langle k\rangle$-manifold. A *$\langle k\rangle$-manifold M* is a smooth manifold with corners so that each codimension-c corner point lies in exactly c facets (closure of a codimension-1 component), equipped with a decomposition of its boundary $\partial M = \cup_{i=1}^{k} \partial_i M$ so that each $\partial_i M$ is a *multifacet* of M (union of disjoint facets), and $\partial_i M \cap \partial_j M$ is a multifacet of $\partial_i M$ and $\partial_j M$, cf. Jänich [1968] and Laures [2000].
3. An associative *composition* map $\mathfrak{M}(y,z) \times \mathfrak{M}(x,y) \hookrightarrow \partial_{\mathrm{gr}(y)-\mathrm{gr}(z)} \mathfrak{M}(x,z) \subset \mathfrak{M}(x,z)$. Setting

   (2-8) $$\mathfrak{M}(i,j) = \coprod_{\substack{x,y \\ \mathrm{gr}(x)=i, \mathrm{gr}(y)=j}} \mathfrak{M}(x,y),$$

   the composition is required to induce an isomorphism of $\langle i-j-2\rangle$-manifolds

   (2-9) $$\partial_{j-k} \mathfrak{M}(i,k) \cong \mathfrak{M}(j,k) \times \mathfrak{M}(i,j).$$

4. *Neat embeddings* $\iota_{i,j}\colon \mathfrak{M}(i,j) \hookrightarrow [0,1)^{i-j-1} \times (-1,1)^{D(i-j)}$ for some large $D \in \mathbb{N}$, namely, smooth embeddings satisfying

   (2-10) $$\iota_{i,k}^{-1}([0,1)^{j-k-1} \times \{0\} \times [0,1)^{i-j-1} \times (-1,1)^{D(i-k)}) = \partial_{j-k} \mathfrak{M}(i,k)$$

   and certain orthogonality conditions near boundaries. These embeddings are required to be coherent with respect to composition. The space of such collections of neat embeddings is $(D-2)$-connected.
5. *Framings* of the normal bundles of $\iota_{i,j}$, also coherent with respect to composition, which give extensions $\bar{\iota}_{i,j}\colon \mathfrak{M}(i,j) \times [-1,1]^{D(i-j)} \hookrightarrow [0,1)^{i-j-1} \times (-1,1)^{D(i-j)}$.

A framed flow category produces a lax linear diagram $P \in \mathfrak{h}\mathscr{T}(S^N)_{\bullet}^{[n+1]}$ with $N = nD$. On objects, set $P(i) = \{x \in \mathscr{C} \mid \mathrm{gr}(x) = i\}$. On morphisms, define the map

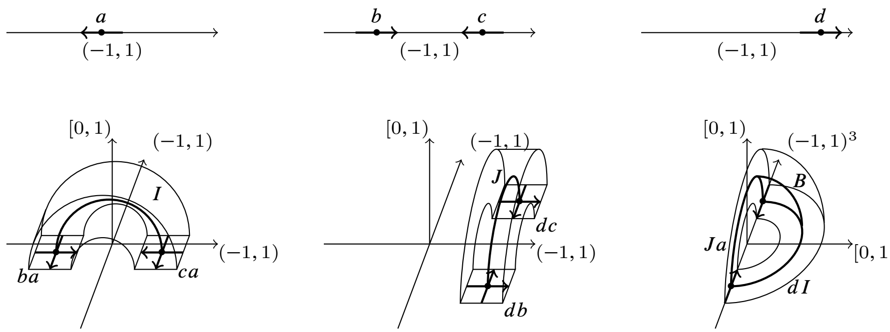


Figure 2.1: **A framed flow category $\mathscr{C}$.** $\mathrm{Ob}(\mathscr{C}) = \{x, y, z, w\}$ *in gradings* $3, 2, 1, 0$*, respectively. The* 0*-dimensional morphism spaces are:* $\mathfrak{M}(x, y) = \{a\}$, $\mathfrak{M}(y, z) = \{b, c\}$*, and* $\mathfrak{M}(z, w) = \{d\}$*, each embedded in* $(-1, 1)$*. The* 1*-dimensional morphism spaces are:* $\mathfrak{M}(x, z) = I$ *(resp.* $\mathfrak{M}(y, w) = J$*) an interval, embedded in* $[0, 1) \times (-1, 1)^2$*, with endpoints* $\{ba, ca\}$ *(resp.* $\{db, dc\}$*) embedded in* $\{0\} \times (-1, 1)^2$ *by the product embedding. The* 2*-dimensional morphism space is a disk* B*, embedded in* $[0, 1)^2 \times (-1, 1)^3$*; it is a* $\langle 2\rangle$*-manifold with boundary decomposed as a union of two arcs,* $\partial_1 B = dI \subset \{0\} \times [0, 1) \times (-1, 1)^3$ *and* $\partial_2 B = Ja \subset [0, 1) \times \{0\} \times (-1, 1)^3$*. Coherent framings of all the normal bundles are represented by the tubular neighborhoods* $\bar{\iota}_{i,j}$*. In the last subfigure,* $(-1, 1)^3$ *is drawn as an interval by projecting to the middle* $(-1, 1)$.

associated to the sequence $m_0 \to m_1 \to \cdots \to m_k$ to be the constant map unless all the arrows are length one. To a sequence of length one arrows, $i \to i - 1 \to \cdots \to j$, associate a map

$$
\begin{aligned}
[0,1]^{i-j-1} \times \bigvee_{x \in P(i)} S^N = [0,1]^{i-j-1} \times \coprod_{x \in P(i)} [-1,1]^{nD}/\partial \\
\to \bigvee_{y \in P(j)} S^N = \coprod_{y \in P(j)} [-1,1]^{nD}/\partial
\end{aligned}
\tag{2-11}
$$

using $\bar{\iota}_{i,j}$ and the Pontryagin-Thom construction.

We can then apply the totalization functor to P to get a cell complex with cells in dimensions $N, N+1, \ldots, N+n$. As Cohen, J. Jones, and Segal [1995] sketch, for a flow category coming from a generic gradient flow of a Morse function, the cell complex produced by the totalization functor is the N^{th} reduced suspension of the Morse cell complex built from the unstable disks of the critical points.

Much of the above data can also be reformulated in the language of S-modules from Pardon [n.d., §4.3]. Let $S[n+1]$ be the (non-symmetric) multicategory with objects pairs

(j,i) of integers with $0 \le j \le i \le n$, unique multimorphisms $(i_0,i_1),(i_1,i_2),\dots,(i_{k-1},i_k) \to (i_0,i_k)$ when $k \ge 1$, and no other multimorphisms (cf. *shape multicategories* from Lawson, Lipshitz, and Sarkar [n.d.(b)]). Let $\underline{\mathsf{Top}}$ be the multicategory of based topological spaces whose multimorphisms $X_1,\dots,X_k \to Y$ are maps $X_1 \wedge \cdots \wedge X_k \to Y$. An S*-module* is a multifunctor $S[n+1] \to \underline{\mathsf{Top}}$.

Given $\mathscr{C}$, define S-modules $\mathbf{S}, \mathbf{J}$ by setting $\mathbf{S}(j,i) = \vee_{y\in P(j)} S^{D(i-j)}$ and $\mathbf{J}(j,i) = (\vee_{x\in P(i)} S^{D(i-j)}) \wedge \mathcal{J}(i,j)$ where $\mathcal{J}$ is the category with objects integers and morphisms $\mathcal{J}(i,j)$ the one-point compactification of $[0,1)^{i-j-1}$ if $i \ge j$ (which is a point if $i = j$) and composition $\mathcal{J}(j,k) \wedge \mathcal{J}(i,j) \to \mathcal{J}(i,k)$ induced by the inclusion map $[0,1)^{j-k-1} \times \{0\} \times [0,1)^{i-j-1} \hookrightarrow \partial[0,1)^{i-k-1}$, cf. Cohen, J. Jones, and Segal [1995, §5]. On a multimorphism $(i_0,i_1),\dots,(i_{k-1},i_k) \to (i_0,i_k)$, $\mathbf{S}$ sends the $(y_0,\dots,y_{k-1}) \in P(i_0) \times \cdots \times P(i_{k-1})$ summand $S^{D(i_1-i_0)} \wedge \cdots \wedge S^{D(i_k-i_{k-1})}$ homeomorphically to the y_0 summand $S^{D(i_k-i_0)}$, and $\mathbf{J}$ sends the $(x_1,\dots,x_k) \in P(i_1)\times\cdots\times P(i_k)$ summand $(S^{D(i_1-i_0)} \wedge \cdots \wedge S^{D(i_k-i_{k-1})}) \wedge (\mathcal{J}(i_1,i_0) \wedge \cdots \wedge \mathcal{J}(i_k,i_{k-1}))$ to the x_k summand $S^{D(i_k-i_0)} \wedge \mathcal{J}(i_k,i_0)$, homeomorphically on the first factor, and using the composition in $\mathcal{J}$ on the second factor. Given a neat embedding of $\mathscr{C}$, we can define another S-module $\mathbf{N}$ by setting $\mathbf{N}(j,i) = \iota^{\nu}_{i,j}$, the Thom space of the normal bundle of $\iota_{i,j}$, if $i > j$. (When $i = j$, $\mathbf{N}(j,j)$ is a point.) On multimorphisms, $\mathbf{N}$ is induced from the inclusion maps $\operatorname{im}(\bar{\iota}_{i_1,i_0})\times\cdots\times\operatorname{im}(\bar{\iota}_{i_k,i_{k-1}}) \to \operatorname{im}(\bar{\iota}_{i_k,i_0})$. The Pontryagin-Thom collapse map is a natural transformation—an S*-module map*—from $\mathbf{J}$ to $\mathbf{N}$, which sends the $x \in P(i)$ summand of $\mathbf{J}(j,i)$ to the Thom-space summand $\cup_{y\in P(j)} \iota^{\nu}_{x,y}$ in $\mathbf{N}(j,i)$. A framing of $\mathscr{C}$ produces another S-module map $\mathbf{N} \to \mathbf{S}$ which sends the summand $\iota^{\nu}_{x,y}$ in $\mathbf{N}(j,i)$ to the y summand of $\mathbf{S}(j,i)$. Composing we get an S-module map $\mathbf{J} \to \mathbf{S}$, which is precisely the data needed to recover a lax diagram in $\mathfrak{h}\mathcal{T}(S^N)^{[n+1]}_{\bullet}$.

Finally, as popularized by Abouzaid, note that since the (smooth) framings of $\iota_{i,j}$ were only used to construct maps $\iota^{\nu}_{i,j} \to \vee_{y\in P(j)} S^{D(i-j)}$, a weaker structure on the flow category—namely, coherent trivializations of the Thom spaces $\iota^{\nu}_{i,j}$ as spherical fibrations—might suffice.

2.5 Speculative digression: matrices of framed cobordisms. Perhaps it would be tidy to reformulate the notion of stably framed flow categories as chain complexes in some category $\mathscr{B}(\mathsf{Cob})$ equipped with a functor $\mathscr{B}(\mathsf{Cob}) \to \mathcal{T}(\mathbb{S})$. It is clear how such a definition would start. Objects in $\mathscr{B}(\mathsf{Cob})$ should be finite sets. By the Pontryagin-Thom construction, a map $\mathbb{S} \to \mathbb{S}$ is determined by a framed 0-manifold; therefore, a morphism in $\mathscr{B}(\mathsf{Cob})$ should be a matrix of framed 0-manifolds. To account for the homotopies in $\mathcal{T}(\mathbb{S})$, $\mathscr{B}(\mathsf{Cob})$ should have higher morphisms. For instance, given two $(T \times S)$-matrices A, B of framed 0-manifolds, a 2-morphism from A to B should be a $(T \times S)$-matrix of framed 1-dimensional cobordisms. Given two such $(T \times S)$-matrices

M, N of framed 1-dimensional cobordisms, a 3-morphism from M to N should be a $(T \times S)$-matrix of framed 2-dimensional cobordisms with corners, and so on. That is, the target category Cob seems to be the extended cobordism category, an (∞, ∞)-category studied, for instance, by Lurie [2009b].

Since matrix multiplication requires only addition and multiplication, the construction $\mathscr{B}(\mathscr{C})$ makes sense for any *rig* or *symmetric bimonoidal* category $\mathscr{C}$ and, presumably, for a rig (∞, ∞)-category, for some suitable definition; and perhaps the framed cobordism category Cob is an example of a rig (∞, ∞)-category. Maybe the Pontryagin-Thom construction gives a functor $\mathscr{B}(\mathsf{Cob}) \to \mathscr{T}(\mathbb{S})$, and that a stably framed flow category is just a functor $[n+1] \to \mathscr{B}(\mathsf{Cob})$.

Rather than pursuing this, we will focus on a tiny piece of Cob, in which all 0-manifolds are framed positively, all 1-dimensional cobordisms are trivially-framed intervals and, more generally, all higher cobordisms are trivially-framed disks. In this case, all of the information is contained in the objects, 1-morphisms, and 2-morphisms, and this tiny piece equals $\mathscr{B}(\mathsf{Sets})$ with Sets being viewed as a rig category via disjoint union and Cartesian product.

2.6 The cube and the Burnside category. The *Burnside category* $\mathscr{B}$ (associated to the trivial group) is the following weak 2-category. The objects are finite sets. The 1-morphisms $\mathrm{Hom}(S, T)$ are $T \times S$ matrices of finite sets; composition is matrix multiplication, using the disjoint union and product of sets in place of $+$ and $\times$ of real numbers. The 2-morphisms are matrices of entrywise bijections between matrices of sets.

(The category $\mathscr{B}$ is denoted $\mathcal{S}_2$ by Hu, D. Kriz, and I. Kriz [2016], and is an example of what they call a *$\star$-category*. The realization procedure below is a concrete analogue of the Elmendorf and Mandell [2006] infinite loop space machine; see also Lawson, Lipshitz, and Sarkar [n.d.(a), §8].)

There is an abelianization functor $\mathsf{Ab}\colon \mathscr{B} \to \mathscr{B}(\mathbb{Z})$ which is the identity on objects and sends a morphism $(A_{t,s})_{s \in S, t \in T}$ to the matrix $(\# A_{t,s})_{s \in S, t \in T} \in \mathbb{Z}^{T \times S}$. We are given a diagram $G \in \mathscr{B}(\mathbb{Z})^{[2]^n}$ which we wish to lift to a diagram $Q \in \mathfrak{h}\mathscr{T}(S^N)^{[2]^n}$. As we will see, it suffices to lift G to a diagram $D\colon [2]^n \to \mathscr{B}$.

Since $\mathscr{B}$ is a weak 2-category, we should first clarify what we mean by a diagram in $\mathscr{B}$. A *strictly unital lax 2-functor*—henceforth just called a lax functor—$D\colon [2]^n \to \mathscr{B}$ consists of the following data:

1. A finite set $F(x) \in \mathscr{B}$ for each $v \in [2]^n$.

2. An $F(v) \times F(u)$-matrix of finite sets $F(u \to v) \in \mathrm{Hom}_{\mathscr{B}}(F(u), F(v))$ for each $u > v \in [2]^n$.

3. A 2-isomorphism $F_{u,v,w}\colon F(v \to w) \circ_1 F(u \to v) \to F(u \to w)$ for each $u > v > w \in [2]^n$ so that for each $u > v > w > z$, $F_{u,w,z} \circ_2 (\mathrm{Id} \circ_1 F_{u,v,w}) = (F_{v,w,z} \circ_1 \mathrm{Id}) \circ_2 F_{u,v,z}$, where $\circ_i$ denotes composition of i-morphisms ($i = 1, 2$).

Next we turn such a lax diagram $D\colon [2]^n \to \mathscr{B}$ into a lax diagram $Q \in \mathfrak{h}\mathscr{T}(S^N)^{[2]^n}$, $N \geq n+1$, satisfying $\widetilde{H}_N \circ Q = \mathsf{Ab} \circ D$. Associate a *box* $B_x = [-1,1]^N$ to each $x \in D(v)$, $v \in [2]^n$. For each $u > v$, let $D(u \to v) = (A_{y,x})_{x \in D(u), y \in D(v)}$ and let $E(u \to v)$ be the space of embeddings $\iota_{u,v} = \{\iota_{u,v,x}\}_{x \in D(u)}$ where

$$\iota_{u,v,x}\colon \coprod_y A_{y,x} \times B_y \hookrightarrow B_x \tag{2-12}$$

whose restriction to each copy of B_y is a sub-box inclusion, i.e., composition of a translation and dilation. The space $E(A)$ is $(N-2)$-connected.

For any such data $\iota_{u,v}$, a collapse map and a fold map give a *box map*

$$\widehat{\iota}_{u,v}\colon \bigvee_{x \in D(u)} S^N = \coprod_{x \in D(u)} B_x/\partial \to \coprod_{\substack{x \in D(u)\\ y \in D(v)\\ a \in A_{y,x}}} B_y/\partial \to \coprod_{y \in D(v)} B_y/\partial = \bigvee_{y \in D(v)} S^N. \tag{2-13}$$

Given $u > v > w$ and data $\iota_{u,v}$, $\iota_{v,w}$, the composition $\widehat{\iota}_{v,w} \circ \widehat{\iota}_{u,v}$ is also a box map corresponding to some induced embedding data.

The construction of the lax diagram $Q \in \mathfrak{h}\mathscr{T}(S^N)^{[2]^n}$ is inductive. On objects, Q agrees with D. For (non-identity) morphisms $u \to v$, choose a box map $Q(u \to v) = \widehat{\iota}_{u,v}\colon \bigvee_{x \in D(u)} S^N \to \bigvee_{y \in D(v)} S^N$ refining $D(u \to v)$. Staying in the space of box maps, the required homotopies exist and are unique up to homotopy because each $E(A)$ is $N-2 \geq n-1$ connected, and there are no sequences of composable morphisms of length $> n-1$. (See Lawson, Lipshitz, and Sarkar [ibid.] for details.)

The above construction closely follows the Pontryagin-Thom procedure from Section 2.4. Indeed, functors from the cube to the Burnside category correspond to certain kinds of flow categories (*cubical* ones), and the realizations in terms of box maps and cubical flow categories agree.

3 Khovanov homology

3.1 The Khovanov cube. Khovanov homology was defined by Khovanov [2000] using the Frobenius algebra $V = H^*(S^2)$. Let $x_- \in H^0(S^2)$ and $x_+ \in H^2(S^2)$ be the positive generators. (Our labeling is opposite Khovanov's convention, as the maps in our cube go from 1 to 0.) Via the equivalence of Frobenius algebras and $(1+1)$-dimensional topological field theories (cf. Abrams [1996]), we can reinterpret V as a functor from the $(1+1)$-dimensional bordism category Cob^{1+1} to $\mathscr{B}(\mathbb{Z})$ that assigns $\{x_+, x_-\}$ to circle, and hence $\prod_{\pi_0(C)}\{x_+, x_-\}$ to a one-manifold C. For $x \in V(C)$, let $\|x\|_+$ (respectively, $\|x\|_-$) denote the number of circles in C labeled x_+ (respectively, x_-) by x. For a cobordism $\Sigma\colon C_1 \to C_0$, the map $V(\Sigma)\colon \mathbb{Z}\langle V(C_1)\rangle = \otimes_{\pi_0(C_1)}\mathbb{Z}\langle x_+, x_-\rangle \to \mathbb{Z}\langle V(C_0)\rangle = \otimes_{\pi_0(C_0)}\mathbb{Z}\langle x_+, x_-\rangle$ is the tensor product of the maps induced by the connected components of Σ; and if $\Sigma\colon C_1 \to C_0$ is a connected, genus-g cobordism, then the (y, x)-entry of the matrix representing the map, $x \in V(C_1)$, $y \in V(C_0)$, is

$$\begin{cases} 1 & \text{if } g = 0,\ \|x\|_+ + \|y\|_- = 1,\\ 2 & \text{if } g = 1,\ \|x\|_+ = \|y\|_- = 0,\\ 0 & \text{otherwise} \end{cases} \tag{3-1}$$

(cf. Bar-Natan [2005] and Hu, D. Kriz, and I. Kriz [2016]).

Now, given a link diagram L with n crossings numbered $c_1, \dots, c_n$, Khovanov [2000] constructs a cubical diagram $G_{Kh} = V \circ \mathbb{L} \in \mathscr{B}(\mathbb{Z})^{[2]^n}$ where $\mathbb{L}\colon [2]^n \to \mathsf{Cob}^{1+1}$ is the *cube of resolutions* (extending Kauffman [1987]) defined as follows. For $v \in [2]^n$, let $\mathbb{L}(v)$ be the complete resolution of the link diagram L formed by resolving the i^{th} crossing c_i $\times$ by the 0-*resolution* $)($ if $v_i = 0$ and by the 1-*resolution* $\asymp$ if $v_i = 1$. For a morphism $u \to v$, $\mathbb{L}(u \to v)$ is the cobordism which is an elementary saddle from the 1-resolution to the 0-resolution near crossings c_i for each i with $u_i > v_i$, and is a product cobordism elsewhere.

The dual of the resulting total complex, shifted by n_-, the number of negatives crossings $\times$ in L, is usually called the Khovanov complex

$$\mathcal{C}^*_{Kh}(L) = \mathsf{Dual}(\mathsf{Tot}(G_{Kh}))[n_-], \tag{3-2}$$

and its cohomology $Kh^*(L)$ the Khovanov homology, which is a link invariant. There is an internal grading, the *quantum grading*, that comes from placing the two symbols x_+ and x_- in two different quantum gradings, and the entire complex decomposes along this grading, so Khovanov homology inherits a second grading $Kh^i(L) = \oplus_j Kh^{i,j}(L)$, and its quantum-graded Euler characteristic

$$\sum_{i,j}(-1)^i q^j \operatorname{rank}(Kh^{i,j}(L)) \tag{3-3}$$

recovers the unnormalized Jones polynomial of L. The quantum grading persists in the space-level refinement but, for brevity, we suppress it.

3.2 The stable homotopy type. Following Section 2.6, to give a space-level refinement of Khovanov homology it suffices to lift G_{Kh} to a lax functor $[2]^n \to \mathscr{B}$.

Hu, D. Kriz, and I. Kriz [2016, §3.2] shows that the TQFT $V\colon \mathsf{Cob}^{1+1} \to \mathscr{B}(\mathbb{Z})$ does not lift to a functor $\mathsf{Cob}^{1+1} \to \mathscr{B}$. However, we may instead work with the *embedded* cobordism category Cob_e^{1+1}, which is a weak 2-category whose objects are closed 1-manifolds embedded in S^2, morphisms are compact cobordisms embedded in $S^2 \times [0,1]$, and 2-morphisms are isotopy classes of isotopies in $S^2 \times [0,1]$ rel boundary. The cube $\mathbb{L}\colon [2]^n \to \mathsf{Cob}^{1+1}$ factors through a functor $\mathbb{L}_e\colon [2]^n \to \mathsf{Cob}_e^{1+1}$. (This functor $\mathbb{L}_e$ is lax, similar to what we had for functors to the Burnside category except without strict unitarity.) So it remains to lift V to a (lax) functor $V_e\colon \mathsf{Cob}_e^{1+1} \to \mathscr{B}$

(3-4)
$$\begin{array}{ccccc} & \overset{\mathbb{L}_e}{\nearrow} & \mathsf{Cob}_e^{1+1} & \overset{V_e}{\dashrightarrow} & \mathscr{B}. \\ [2]^n & & \downarrow & & \downarrow \\ & \underset{\mathbb{L}}{\searrow} & \mathsf{Cob}^{1+1} & \xrightarrow{V} & \mathscr{B}(\mathbb{Z}) \end{array}$$

On an embedded one-manifold C, we must set

(3-5)
$$V_e(C) = \prod_{\pi_0(C)} \{x_+, x_-\}.$$

For an embedded cobordism $\Sigma\colon C_1 \to C_0$ with C_i embedded in $S^2 \times \{i\}$, the matrix $V_e(\Sigma)$ is a tensor product over the connected components of Σ, i.e., if $\Sigma = \amalg_{j=1}^m (\Sigma_j\colon C_{1,j} \to C_{0,j})$ and $(y^j, x^j) \in V_e(C_{0,j}) \times V_e(C_{1,j})$, then the $(\Pi_{j=1}^m y^j, \Pi_{j=1}^m x^j)$ entry of $V_e(\Sigma)$ equals

(3-6)
$$V_e(\Sigma_1)_{y^1,x^1} \times \cdots \times V_e(\Sigma_m)_{y^m,x^m}.$$

And finally, if $\Sigma\colon C_1 \to C_0$ is a connected genus-g cobordism, then for $x \in V_e(C_1)$ and $y \in V_e(C_0)$, the (y,x)-entry of $V_e(\Sigma)$ must be a

(3-7)
$$\begin{cases} \text{1-element set} & \text{if } g = 0,\ \|x\|_+ + \|y\|_- = 1, \\ \text{2-element set} & \text{if } g = 1,\ \|x\|_+ = \|y\|_- = 0, \\ \varnothing & \text{otherwise.} \end{cases}$$

One-element sets do not have any non-trivial automorphisms, so we may set all the one-element sets to $\{\mathrm{pt}\}$. The two-element sets must be chosen carefully: they have to behave naturally under isotopy of cobordisms (the 2-morphisms in Cob_e^{1+1}) and must admit natural isomorphisms $V_e(\Sigma \circ \Sigma') \cong V_e(\Sigma) \circ V_e(\Sigma')$ when composing cobordisms $\Sigma'\colon C_2 \to C_1$ and $\Sigma\colon C_1 \to C_0$.

Decompose $S^2 \times [0,1]$ as a union of two compact 3-manifolds glued along Σ, $A \cup_\Sigma B$. Set $V_e(\Sigma)$ to be the (cardinality two) set of unordered bases $\{\alpha, \beta\}$ for $\ker(H^1(\Sigma) \to$

$H^1(\partial\Sigma)) \cong \mathbb{Z}^2$ so that α (respectively, β) is the restriction of a generator of $\ker(H^1(A) \to H^1(A \cap (S^2 \times \{0,1\}))) \cong \mathbb{Z}$ (respectively, $\ker(H^1(B) \to H^1(B \cap (S^2 \times \{0,1\}))) \cong \mathbb{Z}$), and so that, if we orient Σ as the boundary of A then $\langle \alpha \cup \beta, [\Sigma] \rangle = 1$ (or equivalently, if we orient Σ as the boundary of B then $\langle \beta \cup \alpha, [\Sigma] \rangle = 1$). This assignment is clearly natural.

Given cobordisms $\Sigma' \colon C_2 \to C_1$ and $\Sigma \colon C_1 \to C_0$, we need to construct a natural 2-isomorphism $V_e(\Sigma) \circ V_e(\Sigma') \to V_e(\Sigma \circ \Sigma')$. The only non-trivial case is when Σ and Σ' are genus-0 cobordisms gluing to form a connected, genus-1 cobordism. In that case, letting $x \in V_e(C_2)$ (respectively, $y \in V_e(C_0)$) denote the generator that labels all circles of C_2 by x_- (respectively, all circles of C_0 by x_+), we need to construct a bijection between the (y,x)-entry $M_{y,x}$ of $V_e(\Sigma) \circ V_e(\Sigma')$ and the (y,x)-entry $N_{y,x}$ of $V_e(\Sigma \circ \Sigma')$. Consider an element Z of $M_{y,x}$; Z specifies an element $z \in V(C_1)$. There is a unique circle C in C_1 that is non-separating in $\Sigma \circ \Sigma'$ and is labeled x_+ by z. Choose an orientation o of $\Sigma \circ \Sigma'$, orient C as the boundary of Σ, and let $[C]$ denote the image of C in $H_1(\Sigma \circ \Sigma', \partial(\Sigma \circ \Sigma'))$. Assign to Z the unique basis in $N_{y,x}$ that contains the Poincaré dual of $[C]$. It is easy to check that this map is well-defined, independent of the choice of o, natural, and a bijection.

This concludes the definition of the functor $V_e \colon \mathsf{Cob}_e^{1+1} \to \mathscr{B}$. The spatial lift $Q_{Kh} \in \mathfrak{h}\mathcal{T}(S^N)^{[2]^n}$ is then induced from the composition $V_e \circ \mathbb{L}_e \colon [2]^n \to \mathscr{B}$. Totalization produces a cell complex $\mathsf{Tot}(Q_{Kh})$ with

$$\widetilde{C}^*_{\mathrm{cell}}(\mathsf{Tot}(Q_{Kh}))[N+n_-] = \mathcal{C}^*_{Kh}(L). \tag{3-8}$$

We define the Khovanov spectrum $\mathcal{X}_{Kh}(L)$ to be the formal $(N+n_-)^{\text{th}}$ desuspension of $\mathsf{Tot}(Q_{Kh})$. The stable homotopy type of $\mathcal{X}_{Kh}(L)$ is a link invariant; see Lipshitz and Sarkar [2014a], Hu, D. Kriz, and I. Kriz [2016], and Lawson, Lipshitz, and Sarkar [n.d.(a)]. The spectrum decomposes as a wedge sum over quantum gradings, $\mathcal{X}_{Kh}(L) = \bigvee_j \mathcal{X}^j_{Kh}(L)$. There is also a reduced version of the Khovanov stable homotopy type, $\widetilde{\mathcal{X}}_{Kh}(L)$, refining the reduced Khovanov chain complex.

3.3 Properties and applications. In order to apply the Khovanov homotopy type to knot theory, one needs to extract some concrete information from it beyond Khovanov homology. Doing so, one encounters three difficulties:

☹1. The number of vertices of the Khovanov cube is 2^n, where n is the number of crossings of L, so the number of cells in the CW complex $\mathcal{X}_{Kh}(L)$ grows at least that fast. So, direct computation must be by computer, and for relatively low crossing number links.

☹2. For low crossing number links, $Kh^{i,j}(L)$ is supported near the diagonal $2i - j = \sigma(L)$, so each $\mathcal{X}^j_{Kh}(L)$ has nontrivial homology only in a small number of adjacent gradings, and these $Kh^{i,j}(L)$ have no p-torsion for $p > 2$. If X is a spectrum so that $\widetilde{H}^i(X)$ is nontrivial only for $i \in \{k, k+1\}$ then the homotopy type of X is determined by $\widetilde{H}^*(X)$, while if $\widetilde{H}^*(X)$ is nontrivial in only three adjacent gradings and has no p-torsion ($p > 2$) then the homotopy type of X is determined by $\widetilde{H}^*(X)$ and the Steenrod operations Sq^1 and Sq^2 (see Baues [1995, Theorems 11.2, 11.7]).

☹3. There are no known formulas for most algebro-topological invariants of a CW complex. (The situation is a bit better for simplicial complexes.)

Lipshitz and Sarkar [2014c] found an explicit formula for the operation $\mathrm{Sq}^2\colon Kh^{i,j}(L;\mathbb{F}_2) \to Kh^{i+2,j}(L;\mathbb{F}_2)$. The operation Sq^1 is the Bockstein, and hence easy to compute. Using these, one can determine the spectra $\mathcal{X}^j_{Kh}(L)$ for all prime links up to 11 crossings. All these spectra are wedge sums of (de)suspensions of 6 basic pieces (cf. ☹2), and all possible basic pieces except $\mathbb{C}P^2$ occur (see ?1). The first knot for which $\mathcal{X}^{i,j}_{Kh}(K)$ is not a Moore space is also the first non-alternating knot: $T(3,4)$. Extending these computations:

Theorem 1 (Seed [n.d.]). *There are pairs of knots with isomorphic Khovanov cohomologies but non-homotopy equivalent Khovanov spectra.*

The first such pair is 11^n_{70} and 13^n_{2566}. D. Jones, Lobb, and Schütz [n.d.(b)] introduced moves and simplifications allowing them to give a by-hand computation of Sq^2 for $T(3,4)$ and some other knots.

Theorem 2 (Lawson, Lipshitz, and Sarkar [n.d.(a)]). *Given links L, L', $\mathcal{X}_{Kh}(L \amalg L') \simeq \mathcal{X}_{Kh}(L) \wedge \mathcal{X}_{Kh}(L')$ and, if L and L' are based, $\widetilde{\mathcal{X}}_{Kh}(L\#L') \simeq \widetilde{\mathcal{X}}_{Kh}(L) \wedge \widetilde{\mathcal{X}}_{Kh}(L')$ and $\mathcal{X}_{Kh}(L\#L') \simeq \mathcal{X}_{Kh}(L) \wedge_{\mathcal{X}_{Kh}(U)} \mathcal{X}_{Kh}(L')$. Finally, if $m(L)$ is the mirror of L then $\mathcal{X}_{Kh}(m(L))$ is the Spanier-Whitehead dual to $\mathcal{X}_{Kh}(L)$.*

Corollary 3.1. *For any integer k there is a knot K so that the operation $\mathrm{Sq}^k\colon Kh^{*,*}(K) \to Kh^{*+k,*}(K)$ is nontrivial. (Compare ?3.)*

Proof. Choose a knot K_0 so that in some quantum grading, $\widetilde{Kh}(K_0)$ has 2-torsion but $\widetilde{Kh}(K_0;\mathbb{F}_2)$ has vanishing Sq^i for $i > 1$. (For instance, $K_0 = 13^n_{3663}$ works, Shumakovitch [2014].) Let $K = \overbrace{K_0\#\cdots\#K_0}^{k}$. By the Cartan formula, $\mathrm{Sq}^k(\alpha) \neq 0$ for some $\alpha \in \widetilde{Kh}(K;\mathbb{F}_2)$. The short exact sequence

$$0 \to \widetilde{Kh}(K;\mathbb{F}_2) \to Kh(K;\mathbb{F}_2) \to \widetilde{Kh}(K;\mathbb{F}_2) \to 0$$

from Rasmussen [2005, §4.3] is induced by a cofiber sequence of Khovanov spectra from Lipshitz and Sarkar [2014a, §8], so if $\beta \in Kh(K;\mathbb{F}_2)$ is any preimage of α then by naturality, $\mathrm{Sq}^k(\beta) \neq 0$, as well. □

Plamenevskaya [2006] defined an invariant of links L in S^3 transverse to the standard contact structure, as an element of the Khovanov homology of L.

Theorem 3 (Lipshitz, Ng, and Sarkar [2015]). *Given a transverse link L in S^3 there is a well-defined cohomotopy class of $\mathcal{X}_{Kh}(L)$ lifting Plamenevskaya's invariant.*

While Lipshitz, Ng, and Sarkar [ibid.] show that Plamenevskaya's class is known to be invariant under flypes, the homotopical refinement is not presently known to be. It remains open whether either invariant is effective (i.e., stronger than the self-linking number).

The Steenrod squares on Khovanov homology was used by Lipshitz and Sarkar [2014b] to tweak the concordance invariant and slice-genus bound s by Rasmussen [2010] to give potentially new concordance invariants and slice genus bounds. In the simplest case, Sq^2, these concordance invariants are, indeed, different from Rasmussen's invariants. They can be used to give some new results on the 4-ball genus for certain families of knots, see Lawson, Lipshitz, and Sarkar [n.d.(a)]. More striking, Feller, Lewark, and Lobb [n.d.] used these operations to resolve whether certain knots are *squeezed*, i.e., occur in a minimal-genus cobordism between positive and negative torus knots.

In a different direction, the Khovanov homotopy type admits a number of extensions. Lobb, Orson, and Schütz [2017] and, independently, Willis [n.d.] proved that the Khovanov homotopy type stabilizes under adding twists, and used this to extend it to a colored Khovanov stable homotopy type; further stabilization results were proved by Willis [ibid.] and Islambouli and Willis [n.d.]. D. Jones, Lobb, and Schütz [n.d.(a)] proposed a homotopical refinement of the $\mathfrak{sl}_n$ Khovanov-Rozansky homology for a large class of knots and there is also work in progress in this direction by Hu, I. Kriz, and Somberg [n.d.]. Sarkar, Scaduto, and Stoffregen [n.d.] gave a homotopical refinement of the odd Khovanov homology of Ozsváth, Rasmussen, and Szabó [2013].

The construction of the functor V_e is natural enough that it was used by Lawson, Lipshitz, and Sarkar [n.d.(b)] to give a space-level refinement of the arc algebras and tangle invariants from Khovanov [2002]. In the refinement, the arc algebras are replaced by ring spectra (or, if one prefers, spectral categories), and the tangle invariants by module spectra.

3.4 Speculation.

We conclude with some open questions:

Ⓠ1. Does $\mathbb{C}P^2$ occur as a wedge summand of the Khovanov spectrum associated to some link? (Cf. Section 3.3.) More generally, are there non-obvious restrictions on the spectra which occur in the Khovanov homotopy types?

Ⓠ2. Is the obstruction to amphichirality coming from the Khovanov spectrum stronger than the obstruction coming from Khovanov homology? Presumably the answer is "yes," but verifying this might require interesting new computational techniques.

?3. Are there prime knots with arbitrarily high Steenrod squares? Other power operations? Again, we expect that the answer is "yes."

?4. How can one compute Steenrod operations, or stable homotopy invariants beyond homology, from a flow category? (Compare Lipshitz and Sarkar [2014c].)

?5. Is the refined Plamenevskaya invariant from Lipshitz, Ng, and Sarkar [2015] effective? Alternatively, is it invariant under negative flypes / SZ moves?

?6. Is there a well-defined homotopy class of maps of Khovanov spectra associated to an isotopy class of link cobordisms $\Sigma \subset [0,1] \times \mathbb{R}^3$? Given such a cobordism Σ in general position with respect to projection to $[0,1]$, there is an associated map, but it is not known if this map is an isotopy invariant. More generally, one could hope to associate an $(\infty, 1)$-functor from a quasicategory of links and embedded cobordisms to a quasicategory of spectra, allowing one to study families of cobordisms. If not, this is a sense in which Khovanov homotopy, or perhaps homology, is *un*natural. Applications of these cobordism maps would also be interesting (cf. Swann [2010]).

?7. If analytic difficulties are resolved, applying the Cohen-Jones-Segal construction to the symplectic Khovanov homology of Seidel and Smith [2006] should also give a Khovanov spectrum. Is that symplectic Khovanov spectrum homotopy equivalent to the combinatorial Khovanov spectrum? (Cf. Abouzaid and Smith [n.d.].)

?8. The (symplectic) Khovanov complex admits, in some sense, an $O(2)$-action, cf. Manolescu [2006], Seidel and Smith [2010], Hendricks, Lipshitz, and Sarkar [n.d.], and Sarkar, Seed, and Szabó [2017]. Does the Khovanov stable homotopy type?

?9. Is there a homotopical refinement of the Lee [2005] or Bar-Natan [2005] deformation of Khovanov homology? Perhaps no genuine spectrum exists, but one can hope to find a lift of the theory to a module over *ku* or *ko* or another ring spectrum (cf. Cohen [2009]). Exactly how far one can lift the complex might be predicted by the polarization class of a partial compactification of the symplectic Khovanov setting from Seidel and Smith [2006].

?10. Can one make the discussion in Section 2.5 precise? Are there other rig (or ∞-rig) categories, beyond Sets, useful in refining chain complexes in categorification or Floer theory to get modules over appropriate ring spectra?

?11. Is there an intrinsic, diagram-free description of $\mathcal{X}_{Kh}(K)$ or, for that matter, for Khovanov homology or the Jones polynomial?

References

Mohammed Abouzaid and Thomas Kragh (2016). "On the immersion classes of nearby Lagrangians". *J. Topol.* 9.1, pp. 232–244. MR: 3465849 (cit. on p. 1172).

Mohammed Abouzaid and Ivan Smith (n.d.). "Khovanov homology from Floer cohomology". arXiv: 1504.01230 (cit. on pp. 1172, 1187).

Lowell Abrams (1996). "Two-dimensional topological quantum field theories and Frobenius algebras". *J. Knot Theory Ramifications* 5.5, pp. 569–587. MR: 1414088 (cit. on p. 1182).

Dror Bar-Natan (2002). "On Khovanov's categorification of the Jones polynomial". *Algebr. Geom. Topol.* 2, 337–370 (electronic). MR: 1917056 (cit. on p. 1172).

– (2005). "Khovanov's homology for tangles and cobordisms". *Geom. Topol.* 9, pp. 1443–1499. MR: 2174270 (cit. on pp. 1182, 1187).

Stefan Bauer (2004). "A stable cohomotopy refinement of Seiberg-Witten invariants. II". *Invent. Math.* 155.1, pp. 21–40. MR: 2025299 (cit. on p. 1172).

Stefan Bauer and Mikio Furuta (2004). "A stable cohomotopy refinement of Seiberg-Witten invariants. I". *Invent. Math.* 155.1, pp. 1–19. MR: 2025298 (cit. on p. 1172).

Hans Joachim Baues (1995). "Homotopy types". In: *Handbook of algebraic topology*. Amsterdam: North-Holland, pp. 1–72. MR: 1361886 (cit. on p. 1185).

Raoul Bott (1980). "Marston Morse and his mathematical works". *Bull. Amer. Math. Soc. (N.S.)* 3.3, pp. 907–950. MR: 585177 (cit. on p. 1171).

Aldridge Bousfield and Daniël Kan (1972). *Homotopy limits, completions and localizations*. Lecture Notes in Mathematics, Vol. 304. Berlin: Springer-Verlag, pp. v+348. MR: 0365573 (cit. on p. 1174).

Gunnar Carlsson (1981). "A counterexample to a conjecture of Steenrod". *Invent. Math.* 64.1, pp. 171–174. MR: 621775 (cit. on p. 1173).

Ralph Cohen (2009). "Floer homotopy theory, realizing chain complexes by module spectra, and manifolds with corners". In: *Algebraic topology*. Vol. 4. Abel Symp. Springer, Berlin, pp. 39–59. MR: 2597734 (cit. on pp. 1172, 1187).

– (2010). "The Floer homotopy type of the cotangent bundle". *Pure Appl. Math. Q.* 6.2, Special Issue: In honor of Michael Atiyah and Isadore Singer, pp. 391–438. MR: 2761853 (cit. on p. 1172).

Ralph Cohen, John Jones, and Graeme Segal (1995). "Floer's infinite-dimensional Morse theory and homotopy theory". In: *The Floer memorial volume*. Vol. 133. Progr. Math. Basel: Birkhäuser, pp. 297–325. MR: 1362832 (cit. on pp. 1172, 1175, 1177–1179).

Jean-Marc Cordier (1982). "Sur la notion de diagramme homotopiquement cohérent". *Cahiers Topologie Géom. Différentielle* 23.1. Third Colloquium on Categories, Part VI (Amiens, 1980), pp. 93–112. MR: 648798 (cit. on p. 1176).

Christopher Douglas (n.d.). "Twisted Parametrized Stable Homotopy Theory". arXiv: 0508070 (cit. on p. 1172).

Anthony Elmendorf and Michael Mandell (2006). "Rings, modules, and algebras in infinite loop space theory". *Adv. Math.* 205.1, pp. 163–228. MR: 2254311 (cit. on p. 1180).

Brent Everitt, Robert Lipshitz, Sucharit Sarkar, and Paul Turner (2016). "Khovanov homotopy types and the Dold-Thom functor". *Homology Homotopy Appl.* 18.2, pp. 177–181. MR: 3547241 (cit. on p. 1172).

Brent Everitt and Paul Turner (2014). "The homotopy theory of Khovanov homology". *Algebr. Geom. Topol.* 14.5, pp. 2747–2781. MR: 3276847 (cit. on p. 1172).

Peter Feller, Lukas Lewark, and Andrew Lobb (n.d.). In preparation (cit. on p. 1186).

Andreas Floer (1988a). "An instanton-invariant for 3-manifolds". *Comm. Math. Phys.* 118.2, pp. 215–240. MR: 956166 (cit. on p. 1171).

– (1988b). "Morse theory for Lagrangian intersections". *J. Differential Geom.* 28.3, pp. 513–547. MR: 965228 (cit. on p. 1171).

– (1988c). "The unregularized gradient flow of the symplectic action". *Comm. Pure Appl. Math.* 41.6, pp. 775–813. MR: 948771 (cit. on p. 1171).

Mikio Furuta (2001). "Monopole equation and the $\frac{11}{8}$-conjecture". *Math. Res. Lett.* 8.3, pp. 279–291. MR: 1839478 (cit. on p. 1172).

Kristen Hendricks, Robert Lipshitz, and Sucharit Sarkar (n.d.). "A simplicial construction of G-equivariant Floer homology". arXiv: 1609.09132 (cit. on p. 1187).

Po Hu, Daniel Kriz, and Igor Kriz (2016). "Field theories, stable homotopy theory and Khovanov homology". *Topology Proc.* 48, pp. 327–360 (cit. on pp. 1172, 1180, 1182–1184).

Po Hu, Igor Kriz, and Petr Somberg (n.d.). "Derived representation theory and stable homotopy categorification of sl_k". Currently available at http://www.math.lsa.umich.edu/~ikriz/drt16084.pdf (cit. on p. 1186).

Gabriel Islambouli and Michael Willis (n.d.). "The Khovanov homology of infinite braids". arXiv: 1610.04582 (cit. on p. 1186).

Klaus Jänich (1968). "On the classification of $O(n)$-manifolds". *Math. Ann.* 176, pp. 53–76. MR: 0226674 (cit. on p. 1177).

Dan Jones, Andrew Lobb, and Dirk Schütz (n.d.[a]). "An $\mathfrak{sl}_n$ stable homotopy type for matched diagrams". arXiv: 1506.07725 (cit. on p. 1186).

– (n.d.[b]). "Morse moves in flow categories". arXiv: 1507.03502 (cit. on p. 1185).

Vaughan Jones (1985). "A polynomial invariant for knots via von Neumann algebras". *Bull. Amer. Math. Soc. (N.S.)* 12.1, pp. 103–111. MR: 766964 (cit. on p. 1172).

Louis Kauffman (1987). "State models and the Jones polynomial". *Topology* 26.3, pp. 395–407. MR: 899057 (cit. on p. 1182).

Tirasan Khandhawit (2015a). "A new gauge slice for the relative Bauer-Furuta invariants". *Geom. Topol.* 19.3, pp. 1631–1655. MR: 3352245 (cit. on p. 1172).

Tirasan Khandhawit (2015b). "On the stable Conley index in Hilbert spaces". *J. Fixed Point Theory Appl.* 17.4, pp. 753–773. MR: 3421983 (cit. on p. 1172).

Tirasan Khandhawit, Jianfeng Lin, and Hirofumi Sasahira (n.d.). "Unfolded Seiberg-Witten Floer spectra, I: Definition and invariance". arXiv: 1604.08240 (cit. on p. 1172).

Mikhail Khovanov (2000). "A categorification of the Jones polynomial". *Duke Math. J.* 101.3, pp. 359–426. MR: 1740682 (cit. on pp. 1172, 1181, 1182).

– (2002). "A functor-valued invariant of tangles". *Algebr. Geom. Topol.* 2, 665–741 (electronic). MR: 1928174 (cit. on p. 1186).

Mikhail Khovanov and Lev Rozansky (2008). "Matrix factorizations and link homology". *Fund. Math.* 199.1, pp. 1–91. MR: 2391017 (cit. on p. 1172).

Mikhail Khovanov and Paul Seidel (2002). "Quivers, Floer cohomology, and braid group actions". *J. Amer. Math. Soc.* 15.1, pp. 203–271. MR: 1862802 (cit. on p. 1172).

Thomas Kragh (n.d.). "The Viterbo Transfer as a Map of Spectra". arXiv: 0712.2533 (cit. on p. 1172).

– (2013). "Parametrized ring-spectra and the nearby Lagrangian conjecture". *Geom. Topol.* 17.2. With an appendix by Mohammed Abouzaid, pp. 639–731. MR: 3070514 (cit. on p. 1172).

Peter Kronheimer and Ciprian Manolescu (n.d.). "Periodic Floer pro-spectra from the Seiberg-Witten equations". arXiv: 0203243 (cit. on p. 1172).

Gerd Laures (2000). "On cobordism of manifolds with corners". *Trans. Amer. Math. Soc.* 352.12, 5667–5688 (electronic). MR: 1781277 (cit. on p. 1177).

Tyler Lawson, Robert Lipshitz, and Sucharit Sarkar (n.d.[a]). "Khovanov homotopy type, Burnside category, and products". arXiv: 1505.00213 (cit. on pp. 1172, 1180, 1181, 1184–1186).

– (n.d.[b]). "Khovanov spectra for tangles". arXiv: 1706.02346 (cit. on pp. 1172, 1179, 1186).

– (2017). "The cube and the Burnside category". In: *Categorification in geometry, topology, and physics*. Vol. 684. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 63–85. MR: 3611723 (cit. on p. 1173).

Eun Soo Lee (2005). "An endomorphism of the Khovanov invariant". *Adv. Math.* 197.2, pp. 554–586. MR: 2173845 (cit. on p. 1187).

Robert Lipshitz, Lenhard Ng, and Sucharit Sarkar (2015). "On transverse invariants from Khovanov homology". *Quantum Topol.* 6.3, pp. 475–513. MR: 3392962 (cit. on pp. 1172, 1186, 1187).

Robert Lipshitz and Sucharit Sarkar (2014a). "A Khovanov stable homotopy type". *J. Amer. Math. Soc.* 27.4, pp. 983–1042. MR: 3230817 (cit. on pp. 1172, 1184, 1185).

– (2014b). "A refinement of Rasmussen's s-invariant". *Duke Math. J.* 163.5, pp. 923–952. MR: 3189434 (cit. on pp. 1172, 1186).

– (2014c). "A Steenrod square on Khovanov homology". *J. Topol.* 7.3, pp. 817–848. MR: 3252965 (cit. on pp. 1172, 1185, 1187).

Max Lipyanskiy (n.d.). "Geometric Cycles in Floer Theory". arXiv: 1409.1126 (cit. on p. 1172).

Andrew Lobb, Patrick Orson, and Dirk Schütz (2017). "A Khovanov stable homotopy type for colored links". *Algebr. Geom. Topol.* 17.2, pp. 1261–1281. MR: 3623688 (cit. on p. 1186).

Jacob Lurie (2009a). *Higher topos theory*. Vol. 170. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xviii+925. MR: 2522659 (cit. on p. 1176).

– (2009b). "On the classification of topological field theories". In: *Current developments in mathematics, 2008*. Int. Press, Somerville, MA, pp. 129–280. MR: 2555928 (cit. on p. 1180).

Ciprian Manolescu (2003). "Seiberg-Witten-Floer stable homotopy type of three-manifolds with $b_1 = 0$". *Geom. Topol.* 7, 889–932 (electronic). MR: 2026550 (cit. on p. 1172).

– (2006). "Nilpotent slices, Hilbert schemes, and the Jones polynomial". *Duke Math. J.* 132.2, pp. 311–369. MR: 2219260 (cit. on p. 1187).

– (2007). "Link homology theories from symplectic geometry". *Adv. Math.* 211.1, pp. 363–416. MR: 2313538 (cit. on p. 1172).

John Milnor (1963). *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., pp. vi+153. MR: 0163331 (cit. on p. 1171).

– (1965). *Lectures on the h-cobordism theorem*. Notes by L. Siebenmann and J. Sondow. Princeton University Press, Princeton, N.J., pp. v+116. MR: 0190942 (cit. on p. 1171).

Marston Morse (1925). "Relations between the critical points of a real function of n independent variables". *Trans. Amer. Math. Soc.* 27.3, pp. 345–396. MR: 1501318 (cit. on p. 1171).

– (1930). "The foundations of a theory of the calculus of variations in the large in m-space. II". *Trans. Amer. Math. Soc.* 32.4, pp. 599–631. MR: 1501555 (cit. on p. 1171).

– (1996). *The calculus of variations in the large*. Vol. 18. American Mathematical Society Colloquium Publications. Reprint of the 1932 original. American Mathematical Society, Providence, RI, pp. xii+368. MR: 1451874 (cit. on p. 1171).

Peter Ozsváth, Jacob Rasmussen, and Zoltán Szabó (2013). "Odd Khovanov homology". *Algebr. Geom. Topol.* 13.3, pp. 1465–1488. arXiv: 0710.4300. MR: 3071132 (cit. on p. 1186).

Richard Palais (1963). "Morse theory on Hilbert manifolds". *Topology* 2, pp. 299–340. MR: 0158410 (cit. on p. 1171).

Richard Palais and Stephen Smale (1964). "A generalized Morse theory". *Bull. Amer. Math. Soc.* 70, pp. 165–172. MR: 0158411 (cit. on p. 1171).

John Pardon (n.d.). "Contact homology and virtual fundamental cycles". arXiv: 1508.03873 (cit. on p. 1178).

Olga Plamenevskaya (2006). "Transverse knots and Khovanov homology". *Math. Res. Lett.* 13.4, pp. 571–586. MR: 2250492 (cit. on p. 1186).

Matan Prasma et al. (n.d.). *"Moore space: (Non-)Functoriality of the construction" in nLab*. ncatlab.org/nlab/show/Moore+space (cit. on p. 1173).

Jacob Rasmussen (2005). "Knot polynomials and knot homologies". In: *Geometry and topology of manifolds*. Vol. 47. Fields Inst. Commun. Providence, RI: Amer. Math. Soc., pp. 261–280. MR: 2189938 (cit. on p. 1185).

– (2010). "Khovanov homology and the slice genus". *Invent. Math.* 182.2, pp. 419–447. MR: 2729272 (cit. on p. 1186).

Sucharit Sarkar, Christopher Scaduto, and Matthew Stoffregen (n.d.). "An odd Khovanov homotopy type". arXiv: 1801.06308 (cit. on p. 1186).

Sucharit Sarkar, Cotton Seed, and Zoltán Szabó (2017). "A perturbation of the geometric spectral sequence in Khovanov homology". *Quantum Topol.* 8.3, pp. 571–628. MR: 3692911 (cit. on p. 1187).

Hirofumi Sasahira (n.d.). "Gluing formula for the stable cohomotopy version of Seiberg-Witten invariants along 3-manifolds with $b_1 > 0$". arXiv: 1408.2623 (cit. on p. 1172).

Cotton Seed (n.d.). "Computations of the Lipshitz-Sarkar Steenrod Square on Khovanov Homology". arXiv: 1210.1882 (cit. on p. 1185).

Graeme Segal (1974). "Categories and cohomology theories". *Topology* 13, pp. 293–312. MR: 0353298 (cit. on p. 1174).

Paul Seidel and Ivan Smith (2006). "A link invariant from the symplectic geometry of nilpotent slices". *Duke Math. J.* 134.3, pp. 453–514. MR: 2254624 (cit. on pp. 1172, 1187).

– (2010). "Localization for involutions in Floer cohomology". *Geom. Funct. Anal.* 20.6, pp. 1464–1501. MR: 2739000 (cit. on p. 1187).

Alexander Shumakovitch (2014). "Torsion of Khovanov homology". *Fund. Math.* 225.1, pp. 343–364. MR: 3205577 (cit. on p. 1185).

Stephen Smale (1964). "Morse theory and a non-linear generalization of the Dirichlet problem". *Ann. of Math. (2)* 80, pp. 382–396. MR: 0165539 (cit. on p. 1171).

Jonah Swann (2010). *Relative Khovanov-Jacobsson classes for spanning surfaces*. Thesis (Ph.D.)–Bryn Mawr College. ProQuest LLC, Ann Arbor, MI, p. 119. MR: 2941328 (cit. on p. 1187).

Rainer Vogt (1973). "Homotopy limits and colimits". *Math. Z.* 134, pp. 11–52. MR: 0331376 (cit. on pp. 1174, 1176).

Michael Willis (n.d.). "A Colored Khovanov Homotopy Type And Its Tail For B-Adequate Links". arXiv: 1602.03856 (cit. on p. 1186).

Received 2017-09-25.

Robert Lipshitz
Department of Mathematics, University of Oregon, Eugene, OR 97403
lipshitz@uoregon.edu

Sucharit Sarkar
Department of Mathematics, University of California, Los Angeles, CA 90095
sucharit@math.ucla.edu

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (1193–1210)

HOMOLOGY COBORDISM AND TRIANGULATIONS

CIPRIAN MANOLESCU

Abstract

The study of triangulations on manifolds is closely related to understanding the three-dimensional homology cobordism group. We review here what is known about this group, with an emphasis on the local equivalence methods coming from Pin(2)-equivariant Seiberg-Witten Floer spectra and involutive Heegaard Floer homology.

1 Triangulations of manifolds

A *triangulation* of a topological space X is a homeomorphism $f : |K| \to X$, where $|K|$ is the geometric realization of a simplicial complex K. If X is a smooth manifold, we say that the triangulation is smooth if its restriction to every closed simplex of $|K|$ is a smooth embedding. By the work of Cairns [1935] and Whitehead [1940], every smooth manifold admits a smooth triangulation. Furthermore, this triangulation is unique, up to pre-compositions with piecewise linear (PL) homeomorphisms.

The question of classifying triangulations for *topological* manifolds is much more difficult. Research in this direction was inspired by the following two conjectures.

Hauptvermutung (Steinitz [1908], Tietze [1908]): *Any two triangulations of a space X admit a common refinement (i.e., another triangulation that is a subdivision of both).*

Triangulation Conjecture (based on a remark by Kneser [1926]): *Any topological manifold admits a triangulation.*

Both of these conjectures turned out to be false. The Hauptvermutung was disproved by Milnor [1961], who used Reidemeister torsion to distinguish two triangulations of a space X that is not a manifold. Counterexamples on manifolds came out of the work of Kirby and Siebenmann [1977]. (For a nice survey of the mathematics surrounding the

The author was supported by NSF grant number DMS-1708320.
MSC2010: primary 57R58; secondary 57Q15, 57M27.
Keywords: manifold, gauge theory, Seiberg-Witten, Heegaard Floer.

Hauptvermutung, see Ranicki [1996].) With regard to the triangulation conjecture, counterexamples were shown to exist in dimension 4 by Casson (cf. Akbulut and McCarthy [1990]), and in all dimensions ≥ 5 by Manolescu [2016].

When studying triangulations on manifolds, a natural condition that one can impose is that the link of every vertex is PL homeomorphic to a sphere. Such triangulations are called *combinatorial*, and (up to subdivision) they are equivalent to PL structures on the manifold.

In dimensions ≤ 3, every topological manifold admits a unique smooth and a unique PL structure; cf. Radó [1925] and Moise [1952]. In dimensions ≥ 5, PL structures on topological manifolds were classified in the 1960's. Specifically, building on work of Sullivan [1996] and Casson [1996], Kirby and Siebenmann [1977] proved the following:

- A topological manifold M of dimension $d \geq 5$ admits a PL structure if and only if a certain obstruction class $\Delta(M) \in H^4(M;\mathbb{Z}/2)$ vanishes;
- For every $d \geq 5$, there exists a d-dimensional manifold M such that $\Delta(M) \neq 0$, that is, one without a PL structure;
- If $\Delta(M) = 0$ for a d-dimensional manifold M with $d \geq 5$, then the PL structures on M are classified by elements of $H^3(M;\mathbb{Z}/2)$. (This shows the failure of the Hauptvermutung for manifolds.)

Finally, in dimension four, PL structures are the same as smooth structures, and the classification of smooth structures is an open problem—although much progress has been made using gauge theory, starting with the work of Donaldson [1983]. Note that Freedman [1982] constructed non-smoothable topological four-manifolds, such as the E_8-manifold.

We can also ask about arbitrary triangulations of topological manifolds, not necessarily combinatorial. It is not at all obvious that non-combinatorial triangulations of manifolds exist, but they do.

Example 1.1. Start with a triangulation of a non-trivial homology sphere M^d with $\pi_1(M) \neq 1$; such homology spheres exist in dimensions $d \geq 3$. Take two cones on each simplex, to obtain a triangulation of the suspension ΣM. Repeat the procedure, to get a triangulation of the double suspension $\Sigma^2 M$. By the double suspension theorem of Edwards [2006] and Cannon [1979], the space $\Sigma^2 M$ is homeomorphic to S^{d+2}. However, the link of one of the final cone points is ΣM, which is not even a manifold. Thus, S^{d+2} admits a non-combinatorial triangulation.

Remark 1.2. One can show that any triangulation of a manifold of dimension ≤ 4 is combinatorial.

In general, if we triangulate a d-dimensional manifold, the link of a k-dimensional simplex has the homology of the $(d-k-1)$-dimensional sphere. (However, the link may not be a manifold, as in the example above.) In the 1970's, Galewski and Stern [1980] and Matumoto [1978] developed the theory of triangulations of high-dimensional manifolds by considering homology cobordism relations between the links of simplices. Their theory involves the n-dimensional *homology cobordism group* $\Theta^n_{\mathbb{Z}}$, which we now proceed to define.

Let us first define a d-dimensional *combinatorial homology manifold* M to be a simplicial complex such that the links of k-dimensional simplices have the homology of S^{d-k-1}. We can extend this definition to combinatorial homology manifolds M with boundary, by requiring the links of simplices on the boundary to have the homology of a disk (and so that ∂M is a combinatorial homology manifold). We let

$$\Theta^n_{\mathbb{Z}} = \{Y^n \text{ oriented combinatorial homology manifolds, } H_*(Y) \cong H_*(S^n)\}/\sim$$

where the equivalence relation is given by $Y_0 \sim Y_1 \iff$ there exists a compact, oriented, combinatorial homology manifold W^{n+1} with $\partial W = (-Y_0) \cup Y_1$ and $H_*(W, Y_i; \mathbb{Z}) = 0$. If $Y_0 \sim Y_1$, we say that Y_0 and Y_1 are *homology cobordant*. Summation in $\Theta^n_{\mathbb{Z}}$ is given by connected sum, the standard sphere S^n gives the zero element, and $-[Y]$ is the class of Y with the orientation reversed. This turns $\Theta^n_{\mathbb{Z}}$ into an Abelian group.

It follows from the work of Kervaire [1969] that $\Theta^n_{\mathbb{Z}} = 0$ for $n \neq 3$. On the other hand, the three-dimensional homology cobordism group $\Theta^3_{\mathbb{Z}}$ is nontrivial. To study $\Theta^3_{\mathbb{Z}}$, note that in dimension three, every homology sphere is a manifold. Also, a four-dimensional homology cobordism can be replaced by a PL one, between the same homology spheres, cf. Martin [1973, Theorem A]. Furthermore, in dimensions three and four, the smooth and PL categories are equivalent. This shows that we can define $\Theta^3_{\mathbb{Z}}$ in terms of smooth homology spheres and smooth cobordisms.

The easiest way to see that $\Theta^3_{\mathbb{Z}} \neq 0$ is to consider the *Rokhlin homomorphism*

$$\mu : \Theta^3_{\mathbb{Z}} \to \mathbb{Z}/2, \quad \mu(Y) = \sigma(W)/8 \pmod 2, \tag{1}$$

where W is any compact, smooth, spin 4-manifold with boundary Y, and $\sigma(W)$ denotes the signature of W. For example, the Poincaré sphere P bounds the negative definite plumbing $-E_8$ of signature -8, and therefore has $\mu(P) = 1$. This implies that P is not homology cobordant to S^3, and hence $\Theta^3_{\mathbb{Z}} \neq 0$.

Let us introduce the exact sequence:

$$0 \longrightarrow \ker(\mu) \longrightarrow \Theta^3_{\mathbb{Z}} \xrightarrow{\mu} \mathbb{Z}/2 \longrightarrow 0. \tag{2}$$

We are now ready to state the results of Galewski and Stern [1980], Galewski and Stern [1979] and Matumoto [1978] about triangulations of high-dimensional manifolds.

They mostly parallel those of Kirby-Siebenmann on combinatorial triangulations. One difference is that, when studying arbitrary triangulations, the natural equivalence relation to consider is *concordance*: Two triangulations of the same manifold M are concordant if there exists a triangulation on $M \times [0,1]$ that restricts to the two triangulations on the boundaries $M \times \{0\}$ and $M \times \{1\}$.

- A d-dimensional manifold M (for $d \geq 5$) is triangulable if and only if $\delta(\Delta(M)) = 0 \in H^5(M; \ker(\mu))$. Here, $\Delta(M) \in H^4(M; \mathbb{Z}/2)$ is the Kirby-Siebenmann obstruction to the existence of PL structures, and $\delta : H^4(M; \mathbb{Z}/2) \to H^5(M; \ker(\mu))$ is the Bockstein homomorphism coming from the exact sequence (2).

- There exist non-triangulable manifolds in dimensions ≥ 5 if and only if the exact sequence (2) does not split. (In Manolescu [2016], the author proved that it does not split.)

- If they exist, triangulations on a manifold M of dimension ≥ 5 are classified (up to concordance) by elements in $H^4(M; \ker(\mu))$.

The above results provide an impetus for further studying the group $\Theta^3_{\mathbb{Z}}$, together with the Rokhlin homomorphism.

2 The homology cobordism group

Since $\Theta^3_{\mathbb{Z}}$ can be defined in terms smooth four-dimensional cobordisms, it is not surprising that the tools used to better understand it came from gauge theory. Indeed, beyond the existence of the Rokhlin epimorphism, the first progress was made by Fintushel and Stern [1985], using Yang-Mills theory:

Theorem 2.1 (Fintushel and Stern [ibid.])**.** *The group $\Theta^3_{\mathbb{Z}}$ is infinite. For example, it contains a $\mathbb{Z}$ subgroup, generated by the Poincaré sphere $\Sigma(2,3,5)$.*

Their proof involved associating to a Seifert fibered homology sphere $\Sigma(a_1, \dots, a_k)$ a numerical invariant $R(a_1, \dots, a_k)$, the expected dimension of a certain moduli space of self-dual connections. By combining these methods with Taubes' work on end-periodic four-manifolds (cf. Taubes [1987]), one obtains a stronger result:

Theorem 2.2 (Fintushel and Stern [1990], Furuta [1990])**.** *The group $\Theta^3_{\mathbb{Z}}$ contains a $\mathbb{Z}^\infty$ subgroup. For example, the classes $[\Sigma(2,3,6k-1)]$, $k \geq 1$, are linearly independent in $\Theta^3_{\mathbb{Z}}$.*

When Y is a homology three-sphere, the Yang-Mills equations on $\mathbb{R} \times Y$ were used by Floer [1988] to construct his celebrated *instanton homology*. From the equivariant structure on instanton homology, Frøyshov [2002] defined a homomorphism

$$h : \Theta^3_{\mathbb{Z}} \to \mathbb{Z},$$

with the property that $h(\Sigma(2,3,5)) = 1$ (and therefore h is surjective). This implies the following:

Theorem 2.3 (Frøyshov [ibid.]). *The group $\Theta^3_{\mathbb{Z}}$ has a $\mathbb{Z}$ summand, generated by the Poincaré sphere $P = \Sigma(2,3,5)$.*

Since then, further progress on homology cobordism was made using Seiberg-Witten theory and its symplectic-geometric replacement, Heegaard Floer homology. These will be discussed in Sections 3 and 4, respectively.

In spite of this progress, the following structural questions about $\Theta^3_{\mathbb{Z}}$ remain unanswered:

Questions: *Does $\Theta^3_{\mathbb{Z}}$ have any torsion? Does it have a $\mathbb{Z}^\infty$ summand? Is it in fact $\mathbb{Z}^\infty$?*

We remark that the existence of a $\mathbb{Z}^\infty$ summand could be established by constructing an epimorphism $\Theta^3_{\mathbb{Z}} \to \mathbb{Z}^\infty$. In the context of knot concordance, a result of this type was proved by Hom [2015]: Using knot Floer homology, she showed the existence of a $\mathbb{Z}^\infty$ summand in the smooth knot concordance group generated by topologically slice knots.

3 Seiberg-Witten theory

The Seiberg-Witten equations, introduced in Seiberg and Witten [1994] and Witten [1994] are a prominent tool for studying smooth four-manifolds. They form a system of nonlinear partial differential equations with a $U(1)$ gauge symmetry; the system is elliptic modulo the gauge action. In dimension three, the information coming from these equations can be packaged into an invariant called *Seiberg-Witten Floer homology* (or *monopole Floer homology*). This was defined in full generality, for all three-manifolds, by Kronheimer and Mrowka [2007] in their book. For rational homology spheres, alternate constructions were given in Marcolli and B.-L. Wang [2001], Manolescu [2003], Frøyshov [2010]. Lidman and Manolescu [2016] showed that the definitions in Manolescu [2003] and Kronheimer and Mrowka [2007] are equivalent.

In many settings, the Seiberg-Witten equations can be used as a replacement for the Yang-Mills equations. For example, from the S^1-equivariant structure on Seiberg-Witten Floer homology one can extract an epimorphism

$$\delta : \Theta^3_{\mathbb{Z}} \to \mathbb{Z},$$

and give a new proof of Theorem 2.3; see Frøyshov [2010] and Kronheimer and Mrowka [2007]. It is not known whether δ coincides with the invariant h coming from instanton homology. Note that Frøyshov [2010] and Kronheimer and Mrowka [2007] use the same notation h for the invariant coming from Seiberg-Witten theory; to prevent confusion with the instanton one, we write δ here. We use the normalization that $\delta(P) = 1$ for the Poincaré sphere P.

The construction of Seiberg-Witten Floer homology in Manolescu [2003] actually gives a refined invariant: an S^1-equivariant Floer stable homotopy type, SWF, which can be associated to rational homology spheres equipped with spin^c structures. The definition of SWF was recently generalized to all three-manifolds (in an "unfolded" version) by Khandhawit, J. Lin, and Sasahira [2016].

When the spin^c structure comes from a spin structure, the S^1 symmetry of the Seiberg-Witten equations (given by constant gauge transformations) can be expanded to a symmetry by the group $\mathrm{Pin}(2)$, where

$$\mathrm{Pin}(2) = S^1 \cup jS^1 \subset \mathbb{C} \oplus j\mathbb{C} = \mathbb{H}.$$

As observed in Manolescu [2016], this turns SWF into a $\mathrm{Pin}(2)$-equivariant stable homotopy type, and allows us to define a $\mathrm{Pin}(2)$-equivariant Seiberg-Witten Floer homology. By imitating the construction of the Frøyshov invariant δ in this setting, we obtain three new maps

$$\alpha, \beta, \gamma : \Theta^3_{\mathbb{Z}} \dashrightarrow \mathbb{Z}. \tag{3}$$

These are not homomorphisms (we use the dotted arrow to indicate that), but on the other hand they are related to the Rokhlin homomorphism from (1):

$$\alpha \equiv \beta \equiv \gamma \equiv \mu \pmod 2.$$

Under orientation reversal, the three invariants behave as follows:

$$\alpha(-Y) = -\gamma(Y), \ \beta(-Y) = -\beta(Y).$$

The properties of β suffice to prove the following.

Theorem 3.1 (Manolescu [ibid.]). *There are no 2-torsion elements $[Y] \in \Theta^3_{\mathbb{Z}}$ with $\mu(Y) = 1$. Hence, the short exact sequence (2) does not split and, as a consequence of the work of Galewski and Stern [1980] and Matumoto [1978], non-triangulable manifolds exist in every dimension ≥ 5.*

Indeed, if Y were a homology sphere with $2[Y] = 0 \in \Theta^3_{\mathbb{Z}}$, then Y would be homology cobordant to $-Y$, which would imply that

$$\beta(Y) = \beta(-Y) = -\beta(Y) \ \Rightarrow\ \beta(Y) = 0 \ \Rightarrow\ \mu(Y) = 0.$$

An alternate construction of Pin(2)-equivariant Seiberg-Witten Floer homology, in the spirit of Kronheimer and Mrowka [2007] and applicable to all three-manifolds, was given by F. Lin [2014]. In particular, this gives an alternate proof of Theorem 3.1. Lin's theory was further developed in F. Lin [2015], F. Lin [2016a], F. Lin [2016b], Dai [2016].

The invariants α, β, γ were computed for Seifert fibered spaces by Stoffregen [2015b] and by F. Lin [2015]. One application of their calculations is the following result (a proof of which was also announced earlier by Frøyshov, using instanton homology).

Theorem 3.2 (Frøyshov [n.d.], Stoffregen [2015b], F. Lin [2015]). *There exist homology spheres that are not homology cobordant to any Seifert fibered space.*

This should be contrasted with a result of Myers [1983], which says that every element of $\Theta^3_{\mathbb{Z}}$ can be represented by a hyperbolic three-manifold.

In Stoffregen [2015a], Stoffregen studied the behavior of the invariants α, β, γ under taking connected sums, and used it to give a new proof of the infinite generation of $\Theta^3_{\mathbb{Z}}$. He found a subgroup $\mathbb{Z}^\infty \subset \Theta^3_{\mathbb{Z}}$ generated by the Brieskorn spheres $\Sigma(p, 2p-1, 2p+1)$ for $p \geq 3$ odd. (Compare Theorem 2.2.)

In fact, the information in $\alpha, \beta, \gamma, \delta$, and much more, can be obtained from a stronger invariant, a class in the *local equivalence group* $\mathfrak{LE}$ defined by Stoffregen [2015b]. To define $\mathfrak{LE}$, we first define a *space of type SWF* to be a pointed finite Pin(2)-CW complex X such that

- The S^1-fixed point set X^{S^1} is Pin(2)-homotopy equivalent to $(\tilde{\mathbb{R}}^s)^+$, where $\tilde{\mathbb{R}}$ is the one-dimensional representation of Pin(2) on which S^1 acts trivially and j acts by -1;
- The action of Pin(2) on $X - X^{S^1}$ is free.

The definition is modeled on the properties of the Seiberg-Witten Floer spectra SWF(Y) for homology spheres Y. Any SWF(Y) is the formal (de)suspension of a space of type SWF. The condition on the fixed point set comes from the fact that there is a unique reducible solution to the Seiberg-Witten equations on Y.

The elements of $\mathfrak{LE}$ are equivalence classes $[X]$, where X is a formal (de)suspension of a space of type SWF, and the equivalence relation (called *local equivalence*) is given by: $X_1 \sim X_2 \iff$ there exist Pin(2)-equivariant stable maps

$$\phi : X_1 \to X_2, \quad \psi : X_2 \to X_1,$$

which are both Pin(2)-equivalences on the S^1-fixed point sets. This relation is motivated by the fact that if Y_1 and Y_1 are homology cobordant, then the induced cobordism maps on Seiberg-Witten Floer spectra give a local equivalence between SWF(Y_1) and SWF(Y_2).

We can turn $\mathfrak{LE}$ into an Abelian group, with addition given by smash product, the inverse given by taking the Spanier-Whitehead dual, and the zero element being $[S^0]$. We obtain a group homomorphism

$$\Theta^3_{\mathbb{Z}} \to \mathfrak{LE}, \ \ [Y] \to [\mathrm{SWF}(Y)].$$

The class $[\mathrm{SWF}(Y)] \in \mathfrak{LE}$ encapsulates all known information from Seiberg-Witten theory that is invariant under homology cobordism. The group $\mathfrak{LE}$ is still quite complicated, but there is a simpler version, called the *chain local equivalence group* $\mathfrak{CLE}$, which involves chain complexes rather than stable homotopy types. The elements of $\mathfrak{CLE}$ are modeled on the cellular chain complexes[1] $C^{CW}_*(\mathrm{SWF}(Y);\mathbb{F})$ with coefficients in the field $\mathbb{F} = \mathbb{Z}/2$, viewed as modules over

$$C^{CW}_*(\mathrm{Pin}(2);\mathbb{F}) \cong \mathbb{F}[s,j]/(sj = j^3 s, \ s^2 = 0, j^4 = 1).$$

and divided by an equivalence relation (called chain local equivalence), similar to the one used in the definition of $\mathfrak{LE}$. We have a natural homomorphism

$$\mathfrak{LE} \to \mathfrak{CLE}, \ \ [X] \to [C^{CW}_*(X;\mathbb{F})].$$

To construct interesting maps from $\Theta^3_{\mathbb{Z}}$ to $\mathbb{Z}$, one strategy is to factor them through the groups $\mathfrak{LE}$ or $\mathfrak{CLE}$. Indeed, the Frøyshov homomorphism δ can be obtained this way, by passing from chain complexes to the S^1-equvariant Borel cohomology, which is a module over

$$H^*_{S^1}(pt;\mathbb{F}) = H^*(\mathbb{CP}^\infty;\mathbb{F}) = \mathbb{F}[U], \ \deg(U) = 2.$$

Given the structure of the S^1-fixed point set of $\mathrm{SWF}(Y)$, one can show that $H^*_{S^1}(\mathrm{SWF}(Y);\mathbb{F})$ is the direct sum of an infinite tower $\mathbb{F}[U]$ and an $\mathbb{F}[U]$-torsion part. The invariant $\delta(Y)$ is set to be $1/2$ the minimal grading in the $\mathbb{F}[U]$ tower. The resulting homomorphism δ factors as

$$\Theta^3_{\mathbb{Z}} \longrightarrow \mathfrak{LE} \longrightarrow \mathfrak{CLE} \xrightarrow{\ \delta\ } \mathbb{Z},$$

Here, by a slight abuse of notation, we also used δ to denote the final map from $\mathfrak{CLE}$ to $\mathbb{Z}$.

The maps α, β, γ from (3) are constructed similarly to δ, but using the Pin(2)-equivariant Borel cohomology $H^*_{\mathrm{Pin}(2)}(\mathrm{SWF}(Y);\mathbb{F})$. This is a module over

$$H^*_{\mathrm{Pin}(2)}(pt;\mathbb{F}) = H^*(B\,\mathrm{Pin}(2);\mathbb{F}) = \mathbb{F}[q,v]/(q^3), \ \deg(q) = 1, \ \deg(v) = 4.$$

[1] When applied to SWF, all our chain complexes and homology theories are reduced, but we drop the usual tilde from notation for simplicity.

In this case, if we just consider the $\mathbb{F}[v]$-module structure, we find three infinite towers of the form $\mathbb{F}[v]$, and α, β, γ are the minimal degrees of elements in this towers, suitably renormalized. We can write

$$\Theta^3_{\mathbb{Z}} \longrightarrow \mathfrak{LE} \longrightarrow \mathfrak{CLE} \overset{\alpha,\beta,\gamma}{\dashrightarrow} \mathbb{Z}.$$

Two other numerical invariants $\underline{\delta}, \bar{\delta} : \mathfrak{CLE} \dashrightarrow \mathbb{Z}$ can be obtained by considering the $\mathbb{Z}/4$-equivariant Borel cohomology, where $\mathbb{Z}/4$ is the subgroup

$$\mathbb{Z}/4 = \{1, -1, j, -j\} \subset \mathrm{Pin}(2) = \mathbb{C} \oplus j\mathbb{C}.$$

As shown by Stoffregen [2016], if one considers the Borel homology for other subgroups $G \subset \mathrm{Pin}(2)$, one does not get any information beyond that in $\alpha, \beta, \gamma, \delta, \underline{\delta}$ and $\bar{\delta}$.

However, one can consider other equivariant generalized cohomology theories. For example, there are invariants κ_i, $i \in \{0, 1\}$ coming from $\mathrm{Pin}(2)$-equivariant K-theory (cf. Manolescu [2014] and Furuta and Li [2013]), and κo_i, $i = 0, \dots, 7$, from $\mathrm{Pin}(2)$-equivariant KO-theory (cf. J. Lin [2015]). These factor through $\mathfrak{LE}$, albeit not through $\mathfrak{CLE}$, and have applications to the study of intersection forms of spin four-manifolds with boundary.

In summary, we have a diagram

$$\begin{array}{ccccccc} \Theta^3_{\mathbb{Z}} & \longrightarrow & \mathfrak{LE} & \longrightarrow & \mathfrak{CLE} & \xrightarrow{\ \delta\ } & \mathbb{Z}, \\ & & \Big\downarrow{\scriptstyle \kappa_i, \kappa o_i} & & \Big\downarrow{\scriptstyle \underline{\delta}, \bar{\delta}} & \searrow{\scriptstyle \alpha,\beta,\gamma} & \\ & & \mathbb{Z} & & \mathbb{Z} & & \mathbb{Z} \end{array} \tag{4}$$

Recall that $\mathfrak{CLE}$ was defined using chain complexes with coefficients in $\mathbb{F} = \mathbb{Z}/2$. One could also take coefficients in other fields, say $\mathbb{Q}$ or $\mathbb{Z}/p$ for odd primes p. From the corresponding S^1-equivariant Borel cohomology (with coefficients in a field of characteristic p) one gets homomorphisms

$$\delta_p : \mathfrak{LE} \to \mathbb{Z}.$$

These are different on $\mathfrak{LE}$, but it is not known whether they are different when pre-composed with the map $\Theta^3_{\mathbb{Z}} \to \mathfrak{LE}$. For every homology sphere for which computations are available, the values of δ_p are the same for all p.

On the other hand, Stoffregen [2015b] showed that the information in chain local equivalence (for specific Seifert fibered homology spheres) goes beyond that in the numerical invariants from (4). In fact, using chain local equivalence, he defined an invariant of homology cobordism that takes the form of an Abelian group, called the *connected Seiberg-Witten Floer homology*.

Open problem: *Describe the structure of the groups $\mathcal{LE}$ and $\mathcal{CLE}$, and use it to understand more about $\Theta^3_{\mathbb{Z}}$.*

In particular, it would be interesting to construct more homomorphisms from $\mathcal{LE}$ and $\mathcal{CLE}$ to $\mathbb{Z}$, which could perhaps be used to produce new $\mathbb{Z}$ summands in $\Theta^3_{\mathbb{Z}}$. Of special interest is to construct a lift of the Rokhlin homomorphism to $\mathbb{Z}$, as a homomorphism (rather than just as a map of sets, as is the case with α, β, γ). The existence of such a lift would show that $\Theta^3_{\mathbb{Z}}$ has no torsion with $\mu = 1$, thus strengthening Theorem 3.1. In turn, one can show that this would give a simpler criterion for a high-dimensional manifold to be triangulable: the Galewski-Stern-Matumoto class $\delta(\Delta(M)) \in H^5(M; \ker(\mu))$ could be replaced with an equivalent obstruction in $H^5(M; \mathbb{Z})$.

4 Heegaard Floer homology and its involutive refinement

In a series of papers, Ozsváth and Szabó [2004b], Ozsváth and Szabó [2004a], Ozsváth and Szabó [2006], Ozsváth and Szabó [2003] developed *Heegaard Floer homology*: To every three-manifold Y and spin^c structure $\mathfrak{s}$, they associated invariants

$$\widehat{HF}(Y, \mathfrak{s}),\ HF^+(Y, \mathfrak{s}),\ HF^-(Y, \mathfrak{s}),\ HF^\infty(Y, \mathfrak{s}).$$

These are defined by choosing a pointed Heegaard diagram

$$\mathcal{H} = (\Sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, z)$$

consisting of the Heegaard surface Σ of genus g, two sets of attaching curves $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_g\}$, $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_g\}$, and a basepoint $z \in \Sigma$, away from the attaching curves. The attaching curves describe two handlebodies, which put together should give the three-manifold Y. One then considers the Lagrangians

$$\mathbb{T}_\alpha = \alpha_1 \times \cdots \times \alpha_g,\ \ \mathbb{T}_\beta = \beta_1 \times \cdots \times \beta_g$$

inside the symmetric product $\mathrm{Sym}^g(\Sigma)$. The different flavors ($\widehat{\ }, +, -, \infty$) of Heegaard Floer homology are versions of the Lagrangian Floer homology $HF(\mathbb{T}_\alpha, \mathbb{T}_\beta)$.

The construction of Heegaard Floer homology was inspired by Seiberg-Witten theory: the symmetric product is related to moduli spaces of vortices on Σ. In fact, it has been recently established Kutluhan, Lee, and Taubes [2010], Colin, Ghiggini, and Honda [2012], and Taubes [2010] that Heegaard Floer homology is isomorphic to the monopole Floer homology from Kronheimer and Mrowka [2007]. In view of Lidman and Manolescu [2016], we obtain a relation to the different homologies applied to the Seiberg-Witten Floer spectrum SWF (for rational homology spheres). For example, we have

$$\widehat{HF} \cong H_*(\mathrm{SWF}),\ HF^+ \cong H_*^{S^1}(\mathrm{SWF}).$$

Heegaard Floer homology has had numerous applications to low dimensional topology, and is easier to compute than Seiberg-Witten Floer homology. In fact, it was shown to be algorithmically computable; cf. Sarkar and J. Wang [2010], Lipshitz, Ozsváth, and Thurston [2014], and Manolescu, Ozsváth, and Thurston [2009].

With regard to homology cobordism, in Ozsváth and Szabó [2003] Ozsváth and Szabó defined the *correction terms* $d(Y, \mathfrak{s})$, which are analogues of the Frøyshov invariant δ, and give rise to a homomorphism

$$d : \Theta^3_{\mathbb{Z}} \to \mathbb{Z}.$$

With the usual normalization in Heegaard Floer theory, we have $d = 2\delta$.

One could also ask about recovering $\mathrm{Pin}(2)$-equivariant Seiberg-Witten Floer homology and the invariants α, β, γ using Heegaard Floer homology. For technical reasons (related to higher order naturality), this seems currently out of reach. However, Hendricks and Manolescu [2017] developed *involutive Heegaard Floer homology*, as an analogue of $\mathbb{Z}/4$-equivariant Seiberg-Witten Floer homology, for the subgroup $\mathbb{Z}/4 = \langle j \rangle \subset \mathrm{Pin}(2)$. We start by considering the conjugation symmetry on Heegaard Floer complexes CF° ($\circ \in \{\widehat{\ }, +, -, \infty\}$), coming from interchanging the alpha and beta curves, and reversing the orientation of the Heegaard diagram. When $\mathfrak{s}$ is self-conjugate (i.e., comes from a spin structure), the conjugation symmetry gives rise to an automorphism

$$\iota : CF^\circ(Y, \mathfrak{s}) \to CF^\circ(Y, \mathfrak{s}),$$

which is a homotopy involution, that is, $\iota^2 \sim \mathrm{id}$. We then define the corresponding involutive Heegaard Floer homology as the homology of the mapping cone of $1 + \iota$:

$$HFI^\circ(Y, \mathfrak{s}) = H_*(Cone(CF^\circ(Y) \xrightarrow{(1+\iota)} CF^\circ(Y))).$$

While the usual Heegaard Floer homologies are modules over $H^*_{S^1}(pt) \cong \mathbb{F}[U]$, the involutive versions are modules over $H^*_{\mathbb{Z}/4}(pt) \cong \mathbb{F}[Q, U]/(Q^2)$, with $\deg(U) = -2$, $\deg(Q) = -1$.

Conjecture 4.1. *For every rational homology sphere Y with a self-conjugate* spin^c *structure $\mathfrak{s}$, we have an isomorphism of $\mathbb{F}[Q, U]/(Q^2)$-modules*

$$HFI^+(Y, \mathfrak{s}) \cong H^{\mathbb{Z}/4}_*(\mathrm{SWF}(Y, \mathfrak{s}); \mathbb{F}).$$

From involutive Heegaard Floer homology one can extract invariants $\underline{d}(Y, \mathfrak{s})$, $\bar{d}(Y, \mathfrak{s})$, which are the analogues of (twice) the invariants $\underline{\delta}, \bar{\delta}$ coming from $H^*_{\mathbb{Z}/4}(\mathrm{SWF})$. We get maps

$$\underline{d}, \bar{d} : \Theta^3_{\mathbb{Z}} \dashrightarrow \mathbb{Z}.$$

Involutive Heegaard Floer homology has been computed for various classes of three-manifolds, such as large surgeries on alternating knots (cf. Hendricks and Manolescu [2017]) and the Seifert fibered rational homology spheres $\Sigma(a_1, \dots, a_k)$ (or, more generally, almost-rational plumbings); see Dai and Manolescu [2017]. There is also a connected sum formula for involutive Heegaard Floer homology by Hendricks, Manolescu, and Zemke [2016], and a related connected sum formula for the involutive invariants of knots by Zemke [2017]. The latter had applications to the study of rational cuspidal curves; see Borodzik and Hom [2016].

The calculations of $\underline{d}$ and $\bar{d}$ for the above classes of manifolds (and their connected sums) give more constraints on which 3-manifolds are homology cobordant to each other; see Hendricks and Manolescu [2017], Hendricks, Manolescu, and Zemke [2016], Dai and Stoffregen [2017] for several examples. Furthermore, by imitating Stoffregen's arguments from Stoffregen [2015b], Dai and Manolescu [2017] used *HFI* to give a new proof that $\Theta^3_{\mathbb{Z}}$ has a $\mathbb{Z}^\infty$ subgroup.

The chain local equivalence group $\mathcal{CLE}$ admits an analogue in the involutive context, denoted $\mathfrak{I}$, whose definition is quite simple. Specifically, we define an *ι-complex* to be a pair (C, ι), consisting of

- a $\mathbb{Z}$-graded, finitely generated, free chain complex C over the ring $\mathbb{F}[U]$ ($\deg U = -2$), such that there is a graded isomorphism $U^{-1}H_*(C) \cong \mathbb{F}[U, U^{-1}]$;
- a grading-preserving chain homomorphism $\iota\colon\ C \to C$, such that $\iota^2 \sim \mathrm{id}$.

We say that two ι-complexes (C, ι) and (C', ι') are *locally equivalent* if there exist (grading-preserving) homomorphisms

$$F\colon\ C \to C', \quad G\colon\ C' \to C$$

such that

$$F \circ \iota \simeq \iota' \circ F, \quad G \circ \iota' \simeq \iota \circ G,$$

and F and G induce isomorphisms on $U^{-1}H_*$.

The elements of $\mathfrak{I}$ are the local equivalence classes of ι-complexes, and the multiplication in $\mathfrak{I}$ is given by

$$(C, \iota) * (C', \iota') := (C \otimes_{\mathbb{F}[U]} C',\ \iota \otimes \iota').$$

As shown in Hendricks, Manolescu, and Zemke [2016], there is a homomorphism

$$\Theta^3_{\mathbb{Z}} \to \mathfrak{I}, \quad [Y] \to [(CF^-(Y), \iota)],$$

and the maps $d, \underline{d}, \bar{d}$ factor through $\mathfrak{I}$.

Furthermore, Hendricks, Hom, and Lidman [2018] extract from $\mathfrak{I}$ a new invariant of homology cobordism, the connected Heegaard Floer homology, which is a summand of Heegaard Floer homology.

Open problem: *What is $\mathfrak{I}$ as an Abelian group? Can we use it to say more about $\Theta^3_{\mathbb{Z}}$?*

5 Variations

So far, we only studied homology cobordisms between integer homology spheres. However, one can define homology cobordisms between two arbitrary three-manifolds Y_0 and Y_1, by imposing the same conditions on the cobordism, $H_*(W, Y_i; \mathbb{Z}) = 0, i = 0, 1$. Note that, if Y_0 is homology cobordant to Y_1, then they necessarily have the same homology. The invariants $d, \underline{d}, \bar{d}, \alpha, \beta, \gamma$ admit extensions suitable for studying the existence of homology cobordisms between non-homology spheres; see for example Ozsváth and Szabó [2003, Section 4.2].

We could also weaken the definition of homology cobordism by using homology with coefficients in an Abelian group A different from $\mathbb{Z}$. One gets an *A-homology cobordism group* Θ^3_A, whose elements are A-homology spheres, modulo the relation of A-homology cobordism. Observe, for example, that there are natural maps

$$\Theta^3_{\mathbb{Z}} \to \Theta^3_{\mathbb{Z}/n} \to \Theta^3_{\mathbb{Q}}.$$

Fintushel and Stern [1984] showed that the homology sphere $\Sigma(2,3,7)$ bounds a rational ball, whereas it cannot bound an integer homology ball, because $\mu(\Sigma(2,3,7)) = 1$. This implies that the map $\Theta^3_{\mathbb{Z}} \to \Theta^3_{\mathbb{Q}}$ is not injective. It is also not surjective, and in fact its cokernel is infinitely generated; cf. Kim and Livingston [2014]. In a different direction, Lisca [2007] gave a complete description of the subgroup of $\Theta^3_{\mathbb{Q}}$ generated by lens spaces.

One can also construct other versions of homology cobordism by equipping the three-manifolds with spin^c structures, or self-conjugate spin^c structures. Ozsváth and Szabó did the former in Ozsváth and Szabó [2003], where they defined a spin^c homology cobordism group θ^c, and showed that their correction term gives rise to a homomorphism

$$d : \theta^c \to \mathbb{Q}.$$

The other invariants $\alpha, \beta, \gamma, \underline{d}, \bar{d}$ can be similarly extended to maps from a self-conjugate spin^c homology cobordism group (or, more simply, a spin homology cobordism group) to $\mathbb{Q}$. Moreover, on a $\mathbb{Z}/2$-homology sphere there is a unique self-conjugate spin^c-structure, which we can use to produce maps

$$d, \underline{d}, \bar{d}, \alpha, \beta, \gamma : \Theta^3_{\mathbb{Z}/2} \dashrightarrow \mathbb{Q}.$$

Finally, let us mention that homology cobordism is closely related to knot concordance. Indeed, a concordance between two knots $K_0, K_1 \subset S^3$ gives rise to a homology cobordism between the surgeries $S^3_m(K_0)$ and $S^3_m(K_1)$, for any integer m. It also gives a $\mathbb{Q}$-homology cobordism between the p^n-fold cyclic branched covers $\Sigma_{p^n}(K_0)$ and $\Sigma_{p^n}(K_1)$, for any prime p and $n \geq 1$. Thus, one can get knot concordance invariants from homology cobordism invariants, by applying them to surgeries or branched covers. See Ozsváth and Szabó [2003], Manolescu and Owens [2007], Jabuka [2012], and Hendricks and Manolescu [2017] for examples of this. For a survey of the knot concordance invariants coming from Heegaard Floer homology, we refer to Hom [2017].

Acknowledgments. I would like to thank Jennifer Hom, Charles Livingston, Robert Lipshitz, Andrew Ranicki, Sucharit Sarkar, Matt Stoffregen and Ian Zemke for comments on a previous version of this paper.

References

Selman Akbulut and John D. McCarthy (1990). *Casson's invariant for oriented homology 3-spheres*. Vol. 36. Mathematical Notes. An exposition. Princeton, NJ: Princeton University Press (cit. on p. 1194).

Maciej Borodzik and Jennifer Hom (2016). "Involutive Heegaard Floer homology and rational cuspidal curves". arXiv: 1609.08303 (cit. on p. 1204).

Stewart S. Cairns (1935). "Triangulation of the manifold of class one". *Bull. Amer. Math. Soc.* 41.8, pp. 549–552 (cit. on p. 1193).

James W. Cannon (1979). "Shrinking cell-like decompositions of manifolds. Codimension three". *Ann. of Math. (2)* 110.1, pp. 83–112 (cit. on p. 1194).

A. J. Casson (1996). "Generalisations and applications of block bundles". In: *The Hauptvermutung book*. Vol. 1. *K*-Monogr. Math. Kluwer Acad. Publ., Dordrecht, pp. 33–67 (cit. on p. 1194).

Vincent Colin, Paolo Ghiggini, and Ko Honda (2012). "The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions I". arXiv: 1208.1074 (cit. on p. 1202).

Irving Dai (2016). "On the Pin(2)-equivariant monopole Floer homology of plumbed 3-manifolds". arXiv: 1607.03171 (cit. on p. 1199).

Irving Dai and Ciprian Manolescu (2017). "Involutive Heegaard Floer homology and plumbed three-manifolds". arXiv: 1704.02020 (cit. on p. 1204).

Irving Dai and Matthew Stoffregen (2017). "On homology cobordism and local equivalence between plumbed manifolds". arXiv: 1710.08055 (cit. on p. 1204).

Simon K. Donaldson (1983). "An application of gauge theory to four-dimensional topology". *J. Differential Geom.* 18.2, pp. 279–315 (cit. on p. 1194).
Robert D. Edwards (2006). "Suspensions of homology spheres". arXiv: math/0610573 (cit. on p. 1194).
Ronald Fintushel and Ronald J. Stern (1984). "A μ-invariant one homology 3-sphere that bounds an orientable rational ball". In: *Four-manifold theory (Durham, N.H., 1982)*. Vol. 35. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 265–268 (cit. on p. 1205).
– (1985). "Pseudofree orbifolds". *Ann. of Math. (2)* 122.2, pp. 335–364 (cit. on p. 1196).
– (1990). "Instanton homology of Seifert fibred homology three spheres". *Proc. London Math. Soc. (3)* 61.1, pp. 109–137 (cit. on p. 1196).
Andreas Floer (1988). "An instanton-invariant for 3-manifolds". *Comm. Math. Phys.* 118.2, pp. 215–240 (cit. on p. 1196).
Michael H. Freedman (1982). "The topology of four-dimensional manifolds". *J. Differential Geom.* 17.3, pp. 357–453 (cit. on p. 1194).
Kim A. Frøyshov (n.d.). "Mod 2 instanton Floer homology". in preparation (cit. on p. 1199).
– (2002). "Equivariant aspects of Yang-Mills Floer theory". *Topology* 41.3, pp. 525–552 (cit. on p. 1197).
– (2010). "Monopole Floer homology for rational homology 3-spheres". *Duke Math. J.* 155.3, pp. 519–576 (cit. on pp. 1197, 1198).
Mikio Furuta (1990). "Homology cobordism group of homology 3-spheres". *Invent. Math.* 100.2, pp. 339–355 (cit. on p. 1196).
Mikio Furuta and Tian-Jun Li (2013). "Intersection forms of spin 4-manifolds with boundary". preprint (cit. on p. 1201).
David E. Galewski and Ronald J. Stern (1979). "A universal 5-manifold with respect to simplicial triangulations". In: *Geometric topology (Proc. Georgia Topology Conf., Athens, Ga., 1977)*. New York: Academic Press, pp. 345–350 (cit. on p. 1195).
– (1980). "Classification of simplicial triangulations of topological manifolds". *Ann. of Math. (2)* 111.1, pp. 1–34 (cit. on pp. 1195, 1198).
Kristen Hendricks, Jennifer Hom, and Tye Lidman (2018). "Applications of involutive Heegaard Floer homology". arXiv: 1802.02008 (cit. on p. 1205).
Kristen Hendricks and Ciprian Manolescu (2017). "Involutive Heegaard Floer homology". *Duke Math. J.* 166.7, pp. 1211–1299 (cit. on pp. 1203, 1204, 1206).
Kristen Hendricks, Ciprian Manolescu, and Ian Zemke (2016). "A connected sum formula for involutive Heegaard Floer homology". arXiv: 1607.07499 (cit. on p. 1204).
Jennifer Hom (2015). "An infinite-rank summand of topologically slice knots". *Geom. Topol.* 19.2, pp. 1063–1110 (cit. on p. 1197).
– (2017). "A survey on Heegaard Floer homology and concordance". *J. Knot Theory Ramifications* 26.2, pp. 1740015, 24 (cit. on p. 1206).

Stanislav Jabuka (2012). “Concordance invariants from higher order covers”. *Topology Appl.* 159.10-11, pp. 2694–2710 (cit. on p. 1206).
Michel A. Kervaire (1969). “Smooth homology spheres and their fundamental groups”. *Trans. Amer. Math. Soc.* 144, pp. 67–72 (cit. on p. 1195).
Tirasan Khandhawit, Jianfeng Lin, and Hirofumi Sasahira (2016). “Unfolded Seiberg-Witten Floer spectra, I: Definition and invariance”. arXiv: 1604.08240 (cit. on p. 1198).
Se-Goo Kim and Charles Livingston (2014). “Nonsplittability of the rational homology cobordism group of 3-manifolds”. *Pacific J. Math.* 271.1, pp. 183–211 (cit. on p. 1205).
Robion C. Kirby and Laurence C. Siebenmann (1977). *Foundational essays on topological manifolds, smoothings, and triangulations*. With notes by John Milnor and Michael Atiyah, Annals of Math. Studies, No. 88. Princeton, N.J.: Princeton University Press, pp. vii+355 (cit. on pp. 1193, 1194).
Hellmuth Kneser (1926). “Die Topologie der Mannigfaltigkeiten”. *Jahresbericht der Deut. Math. Verein.* 34, pp. 1–13 (cit. on p. 1193).
Peter B. Kronheimer and Tomasz S. Mrowka (2007). *Monopoles and three-manifolds*. Vol. 10. New Mathematical Monographs. Cambridge: Cambridge University Press (cit. on pp. 1197–1199, 1202).
Cağatay Kutluhan, Yi-Jen Lee, and Clifford H. Taubes (2010). “HF=HM I : Heegaard Floer homology and Seiberg–Witten Floer homology”. arXiv: 1007.1979 (cit. on p. 1202).
Tye Lidman and Ciprian Manolescu (2016). “The equivalence of two Seiberg-Witten Floer homologies”. arXiv: 1603.00582 (cit. on pp. 1197, 1202).
Francesco Lin (2014). “Morse-Bott singularities in monopole Floer homology and the Triangulation conjecture”. arXiv: 1404.4561 (cit. on p. 1199).
– (2015). “The surgery exact triangle in $\mathrm{Pin}(2)$-monopole Floer homology”. arXiv: 1504.01993 (cit. on p. 1199).
– (2016a). “$\mathrm{Pin}(2)$-monopole Floer homology, higher compositions and connected sums”. arXiv: 1605.03137 (cit. on p. 1199).
– (2016b). “Manolescu correction terms and knots in the three-sphere”. arXiv: 1607.05220 (cit. on p. 1199).
Jianfeng Lin (2015). “Pin(2)-equivariant KO-theory and intersection forms of spin 4-manifolds”. *Algebr. Geom. Topol.* 15.2, pp. 863–902 (cit. on p. 1201).
Robert Lipshitz, Peter S. Ozsváth, and Dylan P. Thurston (2014). “Computing $\widehat{HF}$ by factoring mapping classes”. *Geom. Topol.* 18.5, pp. 2547–2681 (cit. on p. 1203).
Paolo Lisca (2007). “Sums of lens spaces bounding rational balls”. *Algebr. Geom. Topol.* 7, pp. 2141–2164 (cit. on p. 1205).
Ciprian Manolescu (2003). “Seiberg–Witten–Floer stable homotopy type of three-manifolds with $b_1 = 0$”. *Geom. Topol.* 7, pp. 889–932 (cit. on pp. 1197, 1198).

– (2014). “On the intersection forms of spin four-manifolds with boundary”. *Math. Ann.* 359.3-4, pp. 695–728 (cit. on p. 1201).

– (2016). “Pin(2)-equivariant Seiberg-Witten Floer homology and the triangulation conjecture”. *J. Amer. Math. Soc.* 29.1, pp. 147–176 (cit. on pp. 1194, 1196, 1198).

Ciprian Manolescu and Brendan Owens (2007). “A concordance invariant from the Floer homology of double branched covers”. *Int. Math. Res. Not. IMRN* 20, Art. ID rnm077, 21 (cit. on p. 1206).

Ciprian Manolescu, Peter S. Ozsváth, and Dylan P. Thurston (2009). “Grid diagrams and Heegaard Floer invariants”. arXiv: `0910.0078` (cit. on p. 1203).

Matilde Marcolli and Bai-Ling Wang (2001). “Equivariant Seiberg-Witten Floer homology”. *Comm. Anal. Geom.* 9.3, pp. 451–639 (cit. on p. 1197).

N. Martin (1973). “On the difference between homology and piecewise-linear bundles”. *J. London Math. Soc. (2)* 6, pp. 197–204 (cit. on p. 1195).

Takao Matumoto (1978). “Triangulation of manifolds”. In: *Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 2*. Proc. Sympos. Pure Math., XXXII. Providence, R.I.: Amer. Math. Soc., pp. 3–6 (cit. on pp. 1195, 1198).

John Milnor (1961). “Two complexes which are homeomorphic but combinatorially distinct”. *Ann. of Math. (2)* 74, pp. 575–590 (cit. on p. 1193).

Edwin E. Moise (1952). “Affine structures in 3-manifolds. V. The triangulation theorem and Hauptvermutung”. *Ann. of Math. (2)* 56, pp. 96–114 (cit. on p. 1194).

Robert Myers (1983). “Homology cobordisms, link concordances, and hyperbolic 3-manifolds”. *Trans. Amer. Math. Soc.* 278.1, pp. 271–288 (cit. on p. 1199).

Peter S. Ozsváth and Zoltán Szabó (2003). “Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary”. *Adv. Math.* 173.2, pp. 179–261 (cit. on pp. 1202, 1203, 1205, 1206).

– (2004a). “Holomorphic disks and three-manifold invariants: properties and applications”. *Ann. of Math. (2)* 159.3, pp. 1159–1245 (cit. on p. 1202).

– (2004b). “Holomorphic disks and topological invariants for closedthree-manifolds”. *Ann. of Math. (2)* 159.3, pp. 1027–1158 (cit. on p. 1202).

– (2006). “Holomorphic triangles and invariants for smooth four-manifolds”. *Adv. Math.* 202.2, pp. 326–400 (cit. on p. 1202).

Tibor Radó (1925). “Über den Begriff der Riemannschen Fläche”. *Acta Sci. Math. (Szeged)* 2, pp. 101–121 (cit. on p. 1194).

Andrew A. Ranicki (1996). “On the Hauptvermutung”. In: *The Hauptvermutung book*. Vol. 1. K-Monogr. Math. Dordrecht: Kluwer Acad. Publ., pp. 3–31 (cit. on p. 1194).

Sucharit Sarkar and Jiajun Wang (2010). “An algorithm for computing some Heegaard Floer homologies”. *Ann. of Math. (2)* 171.2, pp. 1213–1236 (cit. on p. 1203).

Nathan Seiberg and Edward Witten (1994). “Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang-Mills theory”. *Nuclear Phys. B* 426.1, pp. 19–52 (cit. on p. 1197).

Ernst Steinitz (1908). “Beiträge zur Analysis situs”. *Sitz.-Ber. Berl. Math. Ges.* 7, pp. 29–49 (cit. on p. 1193).

Matthew Stoffregen (2015a). “Manolescu invariants of connected sums”. arXiv: 1510.01286 (cit. on p. 1199).

– (2015b). “Pin(2)-equivariant Seiberg-Witten Floer homology of Seifert fibrations”. arXiv: 1505.03234 (cit. on pp. 1199, 1201, 1204).

– (2016). “A Remark on Pin(2)-equivariant Floer homology”. arXiv: 1605.00331 (cit. on p. 1201).

Dennis P. Sullivan (1996). “Triangulating and smoothing homotopy equivalences and homeomorphisms. Geometric Topology Seminar Notes”. In: *The Hauptvermutung book*. Vol. 1. K-Monogr. Math. Dordrecht: Kluwer Acad. Publ., pp. 69–103 (cit. on p. 1194).

Clifford Henry Taubes (1987). “Gauge theory on asymptotically periodic 4-manifolds”. *J. Differential Geom.* 25.3, pp. 363–430 (cit. on p. 1196).

– (2010). “Embedded contact homology and Seiberg-Witten Floer cohomology I”. *Geom. Topol.* 14.5, pp. 2497–2581 (cit. on p. 1202).

Heinrich Tietze (1908). “Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten”. *Monatsh. Math. Phys.* 19.1, pp. 1–118 (cit. on p. 1193).

John Henry C. Whitehead (1940). “On C^1-complexes”. *Ann. of Math. (2)* 41, pp. 809–824 (cit. on p. 1193).

Edward Witten (1994). “Monopoles and Four-Manifolds”. *Math. Res. Lett.* 1, pp. 769–796 (cit. on p. 1197).

Ian Zemke (2017). “Connected sums and involutive knot Floer homology”. arXiv: 1705.01117 (cit. on p. 1204).

Received 2017-10-22.

Ciprian Manolescu
Department of Mathematics
UCLA
520 Portola Plaza
Los Angeles, CA 90095
USA
cm@math.ucla.edu

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (1211–1234)

PROFINITE RIGIDITY

ALAN W. REID

Abstract

We survey recent work on profinite rigidity of residually finite groups.

1 Introduction

It is an old and natural idea to try to distinguish finitely presented groups via their finite quotients, and recently, there has been renewed interest, especially in the light of recent progress in 3-manifold topology, in the question of when the set of finite quotients of a finitely generated residually finite group determines the group up to isomorphism. In more sophisticated terminology, one wants to develop a complete understanding of the circumstances in which finitely generated residually finite groups have isomorphic *profinite completions*. Motivated by this, say that a residually finite group Γ is *profinitely rigid*, if whenever $\widehat{\Delta} \cong \widehat{\Gamma}$, then $\Delta \cong \Gamma$ (see Section 2.2 for definitions and background on profinite completions).

It is the purpose of this article to survey some recent work and progress on profinite rigidity, which is, in part, motivated by Remeslennikov's question (see Question 4.1) on the profinite rigidity of a free group. The perspective taken is that of a low-dimensional topologist, and takes advantage of the recent advances in our understanding of hyperbolic 3-manifolds and their fundamental groups through the pioneering work of Agol [2013] and Wise [2009].

Standing assumption: Throughout the paper all discrete groups considered will be finitely generated and residually finite.

Acknowledgments. Much of the work surveyed here is joint work of the author in several papers with multiple co-authors, and the author would like to sincerely thank his co-authors for their collaborations: M.D.E. Conder, D. B. McReynolds, R. Spitler and H. Wilton. He would like to single out M.R. Bridson for particular thanks in this regard. He would also like to thank M. Boileau, S. Friedl, D. D. Long, A. Lubotzky and P. Zalesskii for many helpful and stimulating conversations on topics related to this work.

The author was supported in part by an NSF grant.
MSC2010: primary 20E18; secondary 57M07.
Keywords: profinite completion, profinite rigidity, 3-manifold group.

2 Preliminaries

We begin by providing some background discussion on profinite groups and profinite completions of discrete groups. We refer the reader to Ribes and Zalesskii [2000] for a more detailed account of the topics covered here.

2.1 Profinite groups. A *directed set* is a partially ordered set I such that for every $i, j \in I$ there exists $k \in I$ such that $k \geq i$ and $k \geq j$. An *inverse system* is a family of sets $\{X_i\}_{\{i\in I\}}$, where I is a directed set, and a family of maps $\phi_{ij} : X_i \to X_j$ whenever $i \geq j$, such that:

- $\phi_{ii} = id_{X_i}$;
- $\phi_{ij}\phi_{jk} = \phi_{ik}$, whenever $i \geq j \geq k$.

Denoting this system by (X_i, ϕ_{ij}, I), the *inverse limit* of the inverse system (X_i, ϕ_{ij}, I) is the set

$$\varprojlim X_i = \{(x_i) \in \prod_{i\in I} X_i | \phi_{ij}(x_i) = x_j,\ \text{whenever } i \geq j\}.$$

If (X_i, ϕ_{ij}, I) is an inverse system of non-empty compact, Hausdorff, totally disconnected topological spaces (resp. topological groups) over the directed set I, then $\varprojlim X_i$ is a non-empty, compact, Hausdorff, totally disconnected topological space (resp. topological group).

In addition, if (X_i, ϕ_{ij}, I) is an inverse system, a subset $J \subset I$ is defined to be *cofinal*, if for each $i \in I$, there exists $j \in J$ with $j \geq i$. If J is cofinal we may form an inverse system (X_j, ϕ_j, J) obtained by omitting those $i \in I \setminus J$. The inverse limit $\varprojlim X_j$ can be identified with the image of $\varprojlim X_i$ under the projection map $\prod_{i\in I} X_i$ onto $\prod_{j\in J} X_j$.

2.2 Profinite completion. Let Γ be a finitely generated group (not necessarily residually finite for this discussion), and let $\mathfrak{N}$ denote the collection of all finite index normal subgroups of Γ. Note that $\mathfrak{N}$ is non-empty as $\Gamma \in \mathfrak{N}$, and we can make $\mathfrak{N}$ into directed set by declaring that

$$\text{for } M, N \in \mathfrak{N},\ M \leq N \text{ whenever } M\ contains\ N.$$

In this case, there are natural epimorphisms $\phi_{NM} : \Gamma/N \to \Gamma/M$, and the inverse limit of the inverse system $(\Gamma/N, \phi_{NM}, \mathfrak{N})$ is denoted $\widehat{\Gamma}$ and defined to be to the *profinite completion* of Γ.

Note that there is a natural map $\iota : \Gamma \to \widehat{\Gamma}$ defined by

$$g \mapsto (gN) \in \varprojlim \Gamma/N,$$

and it is easy to see that ι is injective if and only if Γ is residually finite.

An alternative, perhaps more concrete way of viewing the profinite completion is as follows. If, for each $N \in \mathfrak{N}$, we equip each Γ/N with the discrete topology, then

$\prod\{\Gamma/N : N \in \mathfrak{N}\}$ is a compact space and $\widehat{\Gamma}$ can be identified with $\overline{j(\Gamma)}$ where $j : \Gamma \to \prod\{\Gamma/N : N \in \mathfrak{N}\}$ is the map $g \mapsto (gN)$.

2.3 Profinite Topology. It will also be convenient to recall the profinite topology on a discrete group Γ, its subgroups and the correspondence between the subgroup structure of Γ and $\widehat{\Gamma}$.

The profinite topology on Γ is the topology on Γ in which a base for the open sets is the set of all cosets of normal subgroups of finite index in Γ.

Now given a tower $\mathfrak{T}$ of finite index normal subgroups of Γ:

$$\Gamma > N_1 > N_2 > \dots\dots > N_k > \dots$$

with $\cap N_k = 1$, this can be used to define an inverse system and thereby determines a completion of $\widehat{\Gamma}_{\mathfrak{T}}$ (in which Γ will inject). If the inverse system determined by $\mathfrak{T}$ is cofinal (recall Section 2.1) then the natural homomorphism $\widehat{\Gamma} \to \widehat{\Gamma}_{\mathfrak{T}}$ is an isomorphism. That is to say $\mathfrak{T}$ determines the full profinite topology of Γ.

The following is important in connecting the discrete and profinite worlds (see Ribes and Zalesskii [ibid., p. 3.2.2], where here we use Nikolov and Segal [2007] to replace "open" by "finite index").

Notation. Given a subset X of a profinite group G, we write $\overline{X}$ to denote the closure of X in G.

Proposition 2.1. *If Γ is a finitely generated residually finite group, then there is a one-to-one correspondence between the set $\mathfrak{X}$ of subgroups of Γ that are open in the profinite topology on Γ, and the set $\mathfrak{Y}$ of all finite index subgroups of $\widehat{\Gamma}$.*

Identifying Γ with its image in the completion, this correspondence is given by:

- *For $H \in \mathfrak{X}$, $H \mapsto \overline{H}$.*
- *For $Y \in \mathfrak{Y}$, $Y \mapsto Y \cap \Gamma$.*

If $H, K \in \mathfrak{X}$ and $K < H$ then $[H : K] = [\overline{H} : \overline{K}]$. Moreover, $K \triangleleft H$ if and only if $\overline{K} \triangleleft \overline{H}$, and $\overline{H}/\overline{K} \cong H/K$.

Thus Γ and $\widehat{\Gamma}$ have the same finite quotients. The key result to formalize the precise connection between the collection of finite quotients of Γ and those of $\widehat{\Gamma}$ is the following. This is basically proved in Dixon, Formanek, Poland, and Ribes [1982] (see also Ribes and Zalesskii [2000, pp. 88-89]), the mild difference here, is that we employ Nikolov and Segal [2007] to replace *topological isomorphism* with *isomorphism*. To state this we introduce the following notation:

$$\mathcal{C}(\Gamma) = \{Q : Q \text{ is a finite quotient of } \Gamma\}$$

Theorem 2.2. *Suppose that Γ_1 and Γ_2 are finitely generated abstract groups. Then $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ are isomorphic if and only if $\mathcal{C}(\Gamma_1) = \mathcal{C}(\Gamma_2)$.*

Given this, we make the following definition—this definition is taken, by analogy with the theory of quadratic forms over $\mathbb{Z}$, where two integral quadratic forms can be locally equivalent (i.e. at all places of $\mathbb{Q}$), but not globally equivalent over $\mathbb{Z}$.

Definition 2.3. The genus of a finitely generated residually finite group Γ is: $\mathcal{G}(\Gamma) = \{\Delta : \widehat{\Delta} \cong \widehat{\Gamma}\}$.

In addition, if $\mathcal{P}$ is a class of groups, then we define $\mathcal{G}_{\mathcal{P}}(\Gamma) = \{\Delta \in \mathcal{G}(\Gamma) : \Delta \in \mathcal{P}\}$. For convenience we restate the definition of profinite rigidity.

Definition 2.4. Let Γ be a finitely generated group. Say that Γ is profinitely rigid if $\mathcal{G}(\Gamma) = \{\Gamma\}$.

For convenience we often say that Γ is *profinitely flexible* if it is not profinitely rigid.

In addition when $\Gamma = \pi_1(M)$ where M is a compact 3-manifold we occasionally abuse notation and refer to M as being profinitely rigid or flexible.

The basic questions we are interested in are the following (and also within classes of groups $\mathcal{P}$).

Question 2.5. *Which finitely generated (resp. finitely presented) groups Γ are profinitely rigid (resp. profinitely flexible)?*

Question 2.6. *How large can $|\mathcal{G}(\Gamma)|$ be for finitely generated (resp. finitely presented) groups?*

Question 2.7. *What group theoretic properties are shared by (resp. are different for) groups in the same genus?*

These questions (and ones where the class of groups is restricted) provide the motivation and focus of this article, with particular attention paid to Question 2.5.

2.4 Inducing the full profinite topology. Let Γ be a finitely generated residually finite group and $H < \Gamma$. The profinite topology on Γ determines some pro-topology on H and therefore some completion of H. To understand what happens in certain cases that will be of interest to us, we recall the following. Since we are assuming that Γ is residually finite, H injects into $\widehat{\Gamma}$ and determines a subgroup $\overline{H} \subset \widehat{\Gamma}$. Hence there is a natural epimorphism $\widehat{H} \to \overline{H}$. This need not be injective. For this to be injective (i.e. the full profinite topology is induced on H) it is easy to see that the following needs to hold:

(*) For every subgroup H_1 of finite index in H, there exists a finite index subgroup $\Gamma_1 < \Gamma$ such that $\Gamma_1 \cap H < H_1$.

A important case where the full profinite topology is induced is when the ambient group Γ is *LERF*, the definition of which we recall here. Suppose that Γ is a group and H a subgroup of Γ, then Γ is called H-separable if for every $g \in G \setminus H$, there is a subgroup K of finite index in Γ such that $H \subset K$ but $g \notin K$; equivalently, the intersection of all finite index subgroups in Γ containing H is precisely H. The group Γ is called LERF (or *subgroup separable*) if it is H-separable for every finitely-generated subgroup H, or equivalently, if every finitely-generated subgroup is a closed subset in the profinite topology.

Lemma 2.8. *Let Γ be a finitely-generated group, and H a finitely-generated subgroup of Γ. Suppose that Γ is H_1-separable for every finite index subgroup H_1 in H. Then the profinite topology on Γ induces the full profinite topology on H; that is, the natural map $\widehat{H} \to \overline{H}$ is an isomorphism.*

Proof. Since Γ is H_1 separable, the intersection of all subgroups of finite index in Γ containing H_1 is H_1 itself. From this it easily follows that there exists $\Gamma_1 < \Gamma$ of finite index, so that $\Gamma_1 \cap H = H_1$. The lemma follows from (*) above. □

Immediately from this we deduce.

Corollary 2.9. *Let Γ be a finitely generated group that is LERF. Then if $H < \Gamma$ is finitely generated then the profinite topology on Γ induces the full profinite topology on H; that is, the natural map $\widehat{H} \to \overline{H}$ is an isomorphism.*

3 Two simple examples

We provide two elementary examples that already indicate a level of complexity in trying to understand profinite rigidity and lack thereof. In addition, some consequences of these results and techniques will be helpful in what follows.

Proposition 3.1. *Let Γ be a finitely generated Abelian group, then $\mathcal{G}(\Gamma) = \{\Gamma\}$.*

Proof. Suppose first that $\Delta \in \mathcal{G}(\Gamma)$ and Δ is non-abelian. We may therefore find a commutator $c = [a, b] \in \Delta$ that is non-trivial. Since Δ is residually finite there is a homomorphism $\phi : \Delta \to Q$, with Q finite and $\phi(c) \neq 1$. However, $\Delta \in \mathcal{G}(\Gamma)$, so Q is abelian and therefore $\phi(c) = 1$, a contradiction.

Thus Δ is Abelian, so we can assume that $\Gamma \cong \mathbb{Z}^r \oplus T_1$ and $\Delta \cong \mathbb{Z}^s \oplus T_2$, where T_i $(i = 1, 2)$ are finite Abelian groups. It is easy to see that $r = s$, for if $r > s$ say, we can choose a large prime p such that p does not divide $|T_1||T_2|$, and construct a finite quotient $(\mathbb{Z}/p\mathbb{Z})^r$ that cannot be a quotient of Δ.

In addition if T_1 is not isomorphic to T_2, then some invariant factor appears in T_1 say, but not in T_2. One can then construct a finite abelian group that is a quotient of T_1 (and hence Γ_1) but not of Γ_2. □

Remark 3.2. The proof of Proposition 3.1 also proves the following. Let Γ be a finitely generated group, and suppose that $\Delta \in \mathcal{G}(\Gamma)$. Then $\Gamma^{\mathrm{ab}} \cong \Delta^{\mathrm{ab}}$. In particular $b_1(\Gamma) = b_1(\Delta)$.

Somewhat surprisingly, moving only slightly beyond abelian groups (indeed $\mathbb{Z}$) to groups that are virtually $\mathbb{Z}$, the situation is dramatically different. The following result is due to Baumslag [1974].

Theorem 3.3. *There exists non-isomorphic meta-cyclic groups Γ_1 and Γ_2 for which $\widehat{\Gamma}_1 \cong \widehat{\Gamma}_2$. Both of these groups are virtually $\mathbb{Z}$ and defined as extensions of a fixed finite cyclic group F by $\mathbb{Z}$.*

A more precise form of what Baumslag actually proves in Baumslag [1974] is the following:

Let F be a finite cyclic group with an automorphism of order n, where n is different from 1, 2, 3, 4 and 6. Then there are at least two non-isomorphic cyclic extensions of F, say Γ_1 and Γ_2 with $\widehat{\Gamma}_1 \cong \widehat{\Gamma}_2$.

A beautiful, and useful observation, that is used in the proof that the constructed groups Γ_1 and Γ_2 lie in the same genus is the following that goes back to Hirshon Hirshon [1977]: Suppose that A and B are groups with $A \times \mathbb{Z} \cong B \times \mathbb{Z}$, then $\widehat{A} \cong \widehat{B}$.

Remark 3.4. Moving from meta-cyclic to meta-abelian provides even more striking examples of profinite flexibility. Pickel [1974] constructs finitely presented meta-abelian groups Γ for which $\mathcal{G}(\Gamma)$ is infinite.

4 Profinite rigidity and flexibility in low-dimensions

In connection with Question 2.5 perhaps the most basic case is the following that goes back to Noskov, Remeslennikov, and Romankov [1979, Question 15] and remains open:

Question 4.1. Let F_n be the free group of rank $n \geq 2$. Is F_n profinitely rigid?

The group F_n arises in many guises in low-dimensional topology and affords several natural ways to generalize. In the light of this, natural generalizations of Question 4.1 are the following (which remain open):

Question 4.2. Let Σ_g be a closed orientable surface of genus $g \geq 2$. Is $\pi_1(\Sigma_g)$ profinitely rigid?

As we will discuss in more detail below, profinite rigidity in the setting of 3-manifold groups is different, however, one generalization that we will focus on below is the following question:

Question 4.3. Let M be a complete orientable hyperbolic 3-manifold of finite volume. Is $\pi_1(M)$ profinitely rigid?

In this section we describe some recent progress on Questions 4.1, 4.2 and 4.3, as well as other directions that generalize Question 4.1. However, we begin by recalling some necessary background from the geometry and topology of 3-manifolds.

4.1 Some 3-manifold topology.

For the purposes of this subsection, M will always be a compact connected orientable 3-manifold whose boundary is either empty, or consists of a disjoint union of incompressible tori. The Geometrization Conjecture of

Thurston was established by Perelman (see Morgan and Tian [2014] for a detailed account) and we state what is needed here in a convenient form. We refer the reader to Bonahon [2002] or Thurston [1997] for background on geometric structures on 3-manifolds.

Recall that M is *irreducible* if every embedded 2-sphere in M bounds a 3-ball, and if M is *prime* (i.e. does not decompose as a non-trivial connect sum), then M is irreducible or is covered by $S^2 \times S^1$, in which case M admits a geometric structure modeled on $S^2 \times \mathbb{R}$.

Theorem 4.4. *Let M be an irreducible 3-manifold.*

1. *If $\pi_1(M)$ is finite, then M is covered by S^3.*
2. *If $\pi_1(M)$ is infinite, then M is either:*

   (i) hyperbolic and so arises as $\mathbb{H}^3/\Gamma$ where $\Gamma < \mathrm{PSL}(2,\mathbb{C})$ is a discrete torsion-free subgroup of finite co-volume, or;

   (ii) a Seifert fibered space and has a geometry modeled on $\mathbb{E}^3$, $\mathbb{H}^2 \times \mathbb{R}$, NIL or $\widetilde{SL}_2$, or;

   (iii) a $SOLV$ manifold, or;

   (iv) a manifold that admits a collection of essential tori that decomposes M into pieces that are Seifert fibered spaces with incompressible torus boundary, or have interior admitting a finite volume hyperbolic structure. In this case, we will say that M has a non-trivial JSJ decomposition.

An important well-known consequence of geometrization for us is the following corollary.

Corollary 4.5. *Let M be compact 3-manifold, then $\pi_1(M)$ is residually finite.*

A manifold M that admits a geometric structure modeled on $\mathbb{E}^3$, $S^2 \times \mathbb{R}$, $\mathbb{H}^2 \times \mathbb{R}$, NIL or $SOLV$ all *virtually fiber*. That is to say, given M admitting such a structure then there is a finite cover $M_f \to M$ with M_f constructed as the mapping torus of a surface homeomorphism $f : \Sigma_g \to \Sigma_g$, where $g = 0$ in the case of $S^2 \times \mathbb{R}$, $g = 1$ in the case of $\mathbb{E}^3$, NIL or $SOLV$ and $g > 1$ when the geometry is $\mathbb{H}^2 \times \mathbb{R}$. If M is a compact Seifert fibered space with incompressible torus boundary, then M also virtually fibers. On the other hand, it is known (see Gabai [1986]) that closed manifolds admitting a geometric structure modeled on $\widetilde{SL}_2$ do not virtually fiber.

Regarding virtual fibering of hyperbolic manifolds, a major breakthrough came with Agol's work in Agol [2008], which, taken together with work of Agol [2013] and Wise [2009] (see also Groves and Manning [2017]) leads to the following.

Theorem 4.6 (Virtual fibering)**.** *Let M be a finite volume hyperbolic 3-manifold. Then M has a finite cover that fibers over the circle.*

For manifolds with a non-trivial JSJ decomposition, it was known previously that there were *graph manifolds* (i.e. all pieces in the decomposition are Seifert fibered spaces) that do not virtually fiber Neumann [1997], whilst more recently it was shown in Przytycki and Wise [2018] that *mixed 3-manifolds* (i.e. those in Theorem 4.4 2(iv) that have a decomposition containing a hyperbolic piece) are all virtually fibered.

4.2 Profinite completions of 3-manifold groups after Agol and Wise. The remarkable work of Agol [2013] and Wise [2009] has had significant implications on our understanding of the profinite completion of the fundamental groups of finite volume hyperbolic 3-manifolds. We refer the reader to the excellent book Aschenbrenner, Friedl, and Wilton [2015] for a detailed discussion of the many consequences of Agol [2013] and Wise [2009] for 3-manifold groups. One such concerns LERF (recall Section 2.4). The following result summarizes work of Scott [1978] for Seifert fibered spaces, Agol [2013] and Wise [2009] in the hyperbolic setting, and Sun [2016] who showed that non-geometric irreducible 3-manifolds had non-LERF fundamental group

Theorem 4.7. *Let M be an irreducible 3-manifold (as in Section 4.1). Then $\pi_1(M)$ is LERF if and only if M is geometric (i.e covered by Theorem 4.4 1, 2(i), (ii), (iii)).*

Lemma 2.8 together with Theorem 4.7 yields the following consequence.

Corollary 4.8. *Let M be a finite volume hyperbolic 3-manifold and $H < \pi_1(M)$ a finitely generated subgroup. Then the full profinite topology on H is induced by the profinite topology of $\pi_1(M)$. In particular the closure of H in $\widehat{\pi_1(M)}$ is isomorphic to $\widehat{H}$.*

We now turn to *goodness* in the sense of Serre [1997]. Let G be a profinite group, M a discrete G-module (i.e. an abelian group M equipped with the discrete topology on which G acts continuously) and let $C^n(G, M)$ be the set of all continous maps $G^n \to M$. One defines the coboundary operator $d : C^n(G, M) \to C^{n+1}(G, M)$ in the usual way thereby defining a complex $C^*(G, M)$ whose cohomology groups $H^q(G; M)$ are called the continuous cohomology groups of G with coefficients in M.

Now let Γ be a finitely generated group. Following Serre [ibid.], we say that a group Γ is *good* if for all $q \geq 0$ and for every finite Γ-module M, the homomorphism of cohomology groups

$$H^q(\widehat{\Gamma}; M) \to H^q(\Gamma; M)$$

induced by the natural map $\Gamma \to \widehat{\Gamma}$ is an isomorphism between the cohomology of Γ and the continuous cohomology of $\widehat{\Gamma}$.

Example 4.9. Finitely generated free groups are good.

In general goodness is hard to establish, however, one can establish goodness for a group Γ that is LERF (indeed a weaker version of separability is all that is needed) and in addition has a ”well-controlled splitting of the group” as a graph of groups Grunewald, Jaikin-Zapirain, and Zalesskii [2008]; for example that coming from the virtual special technology Wise [2009]. In addition, a useful criterion for goodness is provided by the next lemma due to Serre [1997, Chapter 1, Section 2.6].

Lemma 4.10. *The group* Γ *is good if there is a short exact sequence*

$$1 \to N \to \Gamma \to H \to 1,$$

such that H *and* N *are good,* N *is finitely-generated, and the cohomology group* $H^q(N, M)$ *is finite for every* q *and every finite* Γ*-module* M.

Coupled with Theorem 4.6 (the virtual fibering theorem) and commensurability invariance of goodness Grunewald, Jaikin-Zapirain, and Zalesskii [2008], this proves that the fundamental groups of all finite volume hyperbolic 3-manifolds are good. Indeed, more is true using Agol [2013] and Wise [2009] (as noticed by Cavendish [2012], see also Reid [2015]):

Theorem 4.11. *Let* M *be a compact* 3*-manifold, then* $\pi_1(M)$ *is good.*

Several notable consequences of this are recorded below.

Corollary 4.12. *Let* M *be a closed irreducible orientable* 3*-manifold, and* N *a compact* 3*-manifold with* $\widehat{\pi_1(M)} \cong \widehat{\pi_1(N)}$*. Then:*

1. $\widehat{\pi_1(M)}$ *is torsion-free.*
2. N *is closed, orientable and can have no summand that has finite fundamental group.*

Proof. Let $\Gamma = \pi_1(M)$ and $\Delta = \pi_1(N)$. Since $\mathrm{cd}(\Gamma) = 3$, $H^3(\Gamma, \mathbb{F}_p) \neq 0$ for every prime p, and $H^q(\Gamma, M) = 0$ for every Γ-module M and every $q > 3$. By goodness, these transfer to the profinite setting in the context of finite modules. It follows from standard results about the cohomology of finite groups, that goodness forces $\widehat{\pi_1(M)}$ to be torsion-free. Hence Δ is also torsion-free, and so N cannot have a summand that has finite fundamental group.

In addition, N must be closed, since $H^3(\Gamma, \mathbb{F}_2) \neq 0$ implies $H^3(\widehat{\Gamma}, \mathbb{F}_2) \neq 0$, and if N is not closed we have, $H^3(\Delta, \mathbb{F}_2) = H^3(\widehat{\Delta}, F_2) = 0$. Orientability follows in a similar fashion using $H^3(\Gamma, \mathbb{F}_p) \neq 0$ for $p \neq 2$. □

Remark 4.13. In Lubotzky [1993], it is shown that there are torsion-free subgroups $\Gamma < \mathrm{SL(n}, \mathbb{Z})$ ($n \geq 3$) of finite index, for which $\widehat{\Gamma}$ contains torsion of all possible orders. It follows that $\mathrm{SL(n}, \mathbb{Z})$ is not good for $n \geq 3$).

4.3 Profinite flexibility of 3-manifold groups. We now describe some recent progress on identifying 3-manifold groups by their profinite completions restricted to the class of 3-manifold groups. To that end let

$$\mathfrak{M} = \{\pi_1(M) : M \text{ is a compact 3-manifold}\}.$$

We note that unlike in the previous subsection M need not be prime, can be non-orientable, may have boundary other than tori and this boundary may be compressible. By capping off 2-sphere boundary components with 3-balls, we can exclude S^2 boundary components (and $\mathbb{RP}^2$ boundary components). Also note that included in $\mathfrak{M}$ are the fundamental groups of non-compact finite volume hyperbolic 3-manifolds where such a manifold is viewed as the interior of a compact 3-manifold with boundary consisting of tori or Klein bottles.

Example 4.14 (Profinitely flexible Seifert fibered spaces)**.** We record a construction of Hempel [2014] that provides examples of closed Seifert fibered spaces M_1 and M_2 that are not homeomorphic but $\widehat{\pi_1(M_1)} \cong \widehat{\pi_1(M_2)}$. This builds on the idea of Baumslag mentioned in Section 3.

Let $f : S \to S$ be a periodic, orientation-preserving homeomorphism of a closed orientable surface S of genus at least 2, and let k be relatively prime to the order of f. Let M_f (resp. M_{f^k}) denote the mapping torus of f (resp. f^k), and let $\Gamma_f = \pi_1(M_f)$ (resp. $\Gamma_{f^k} = \pi_1(M_{f^k})$).

Hempel shows that $\widehat{\Gamma}_f \cong \widehat{\Gamma}_{f^k}$ by proving that $\Gamma_f \times \mathbb{Z} \cong \Gamma_{f^k} \times \mathbb{Z}$ (c.f. the example of Baumslag in Section 3). The proof is elementary group theory, but Hempel also notes that, interestingly, the isomorphism $\Gamma_f \times \mathbb{Z} \cong \Gamma_{f^k} \times \mathbb{Z}$ follows from Kwasik and Rosicki [2004] where it is shown that (in the notation above) $M_f \times S^1 \cong M_{f^k} \times S^1$.

Of course some additional work is needed to prove that the groups are not isomorphic, but in fact typically this is the case as Hempel describes in Hempel [2014]. Note that these examples admit a geometric structure modeled on $\mathbb{H}^2 \times \mathbb{R}$.

More recently it was shown by Wilkes [2017] that the construction of Hempel is the only occasion in which profinite rigidity fails in the closed case (there are also results in the bounded case). More precisely:

Theorem 4.15 (Wilkes)**.** *Let M be a closed Seifert fibered space with infinite fundamental group. Then $\mathcal{G}_{\mathfrak{M}}(\pi_1(M)) = \{\pi_1(M)\}$ unless M is as in Example 4.14 and the failure is precisely given by the construction in Example 4.14. In this case, $\mathcal{G}_{\mathfrak{M}}(\pi_1(M))$ is finite.*

The proof of this relies on some beautiful work of Wilton and Zalesskii [2017a] that remarkably detects geometric structure from finite quotients. We discuss this in more detail below in Section 4.4. but first give some other examples of profinite flexibility in the setting of closed 3-manifolds.

Example 4.16 (Profinitely flexible torus bundles)**.** Profinite flexibility for the fundamental groups of torus bundles admitting a $SOLV$ geometry was studied in detail in Funar [2013]. These torus bundles arise as the mapping torus of a self-homeomorphism $f : T^2 \to T^2$ which can be identified with an element of $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ with $|a+d| > 2$. In Funar [ibid.] it is shown that for any $m \geq 2$ there exist m torus bundles admitting $SOLV$ geometry whose fundamental groups have isomorphic profinite completions but are pairwise non-isomorphic.

A particular pair of examples of such torus bundles are give by the mapping tori of the following homeomorphisms:

$$f_1 = \begin{pmatrix} 188 & 275 \\ 121 & 177 \end{pmatrix} \text{ and } f_1 = \begin{pmatrix} 188 & 11 \\ 3025 & 177 \end{pmatrix}.$$

The methods of proof are very different from that used in Example 4.14. In particular it does not use the ideas in Baumslag's examples in Section 3, using instead, number theoretic techniques arising in understanding "local conjugacy" of matrices in $\mathrm{SL}(2,\mathbb{Z})$. Briefly, the fundamental groups of torus bundles M_f and M_g have isomorphic profinite completions if and only if the cyclic subgroups $< f >, < g > \subset \mathrm{SL}(2,\mathbb{Z})$ are locally conjugate, namely their images modulo m are conjugate in $\mathrm{GL}(2,\mathbb{Z}/\mathrm{m}\mathbb{Z})$, for any positive integer m (see Funar [ibid.]).

Interestingly, as described in Funar [ibid.] the issue of profinite flexibility in this case is related to problems arising from understanding quantum TQFT invariants of the torus bundles.

Example 4.17 (Profinitely flexible 3-manifolds with non-trivial JSJ decomposition)**.** The fundamental groups of the manifolds occurring in Theorem 4.4 2(iv) were investigated in Wilkes [2016]. We will not go into this in any detail here, other than to say that it is shown in Wilkes [ibid.] that there are non-homeomorphic closed graph manifolds whose fundamental groups have isomorphic profinite completions, and that graph manifolds can be distinguished from mixed 3-manifolds by the profinite completion of their fundamental groups. In addition it is shown that if M is a graph manifold that is profinitely flexible, then $|\mathcal{G}_{\mathfrak{M}}(\pi_1(M))| < \infty$.

4.4 Profinite completions of 3-manifold groups and geometric structures. We now turn to the work of Wilton and Zalesskii [2017a,b] that describes a beautiful connection between the existence of a particular geometric structure on a 3-manifold and the profinite completion of its fundamental group. We begin with a mild strengthening of Wilton and Zalesskii [2017a, Theorem 8.4]

Theorem 4.18 (Wilton-Zalesskii)**.** *Let M be a closed orientable 3-manifold with infinite fundamental group admitting one of Thurston's eight geometries and let $N \in \mathfrak{M}$ with $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$. Then N is closed and admits the same geometric structure.*

Proof. This is proved in Wilton and Zalesskii [ibid., Theorem 8.4] with N assumed to be closed, orientable and irreducible. However, the version stated in Theorem 4.18 quickly reduces to this. Briefly, by Theorem 4.11 $\pi_1(M)$ is good, so immediately we have N is closed and orientable by Corollary 4.12.

Furthermore, $\widehat{\pi_1(M)}$ is torsion-free by Corollary 4.12 and so if N is not prime, the summands must all have torsion-free fundamental group. However, in this case we can use the fact that the first L^2-betti number $b_1^{(2)}$ is a profinite invariant by Bridson, Conder, and Reid [2016], and this, together with the work of Lott and Lück [1995] shows that aspherical geometric 3-manifolds have $b_1^{(2)} = 0$, whilst manifolds that are not prime and have torsion-free fundamental group have $b_1^{(2)} \neq 0$. Note that their theorem is stated only for orientable manifolds but this is not a serious problem because, by Lück approximation Lück [1994], if X is a non-orientable compact 3-manifold with infinite fundamental group and $Y \to X$ is its orientable double cover, then $b_1^{(2)}(Y) = 2\, b_1^{(2)}(X)$. We can now use Wilton and Zalesskii [2017a] to complete the proof. □

Given Theorem 4.18, Theorem 4.15, is reduced to the consideration of Seifert fiber spaces. However, the proof still entails some significant work using Bridson, Conder, and Reid [2016] as well as the delicate issue of recovering the euler number of the Seifert fibration from the profinite completion.

In the context of hyperbolic manifolds, a corollary of Theorem 4.18 that is worth recording is.

Corollary 4.19. *Let M be a closed orientable hyperbolic 3-manifold and $N \in \mathfrak{M}$ with $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$, then N is closed orientable and hyperbolic.*

More recently Wilton and Zalesskii [2017b] have established a cusped version of this result, namely.

Theorem 4.20. *Let M be a finite volume non-compact orientable hyperbolic 3-manifold and $N \in \mathfrak{M}$ with $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$, then N is a finite volume non-compact orientable hyperbolic 3-manifold.*

The proofs of Theorems 4.18 and 4.20 also use the work of Agol and Wise, as well as crucially using "nice" actions of profinite groups on profinite trees which are transferred from the discrete setting using LERF and other parts of the virtual special technology of Wise [2009] (see Wilton and Zalesskii [2017a] and Wilton and Zalesskii [2017b] for details).

Actually what is really at the heart of Corollary 4.19 is a profinite analogue of the Hyperbolization Theorem, which asserts that M is hyperbolic if and only if $\pi_1(M)$ does not contain a copy of $\mathbb{Z} \oplus \mathbb{Z}$. The main part of the proof of Corollary 4.19 is to show that if M is a closed hyperbolic 3-manifold, then $\widehat{\pi_1(M)}$ does not contain a subgroup isomorphic to $\widehat{\mathbb{Z}} \oplus \widehat{\mathbb{Z}}$.

Remark 4.21. One might wonder about the extent to which the full profinite completion of the fundamental group of a hyperbolic 3-manifold is actually needed to distinguish the fundamental group. With that in mind, it is easy to give infinitely many examples of links $L \subset S^3$ (so-called homology boundary links) with hyperbolic complement for which $\pi_1(S^3 \setminus L)$ all have the same pro-p completion (namely the free pro-p group of rank 2) for all primes p, see Bridson and Reid [2015a, Section 8.4] for an explicit example.

4.5 Profinite rigidity amongst 3-manifold groups. We now turn to the issue of profinite rigidity. Given the discussion in Section 4.3 about the failure of profinite rigidity (even amongst 3-manifold groups) for Seifert fibered spaces, torus bundles admitting $SOLV$ geometry, and manifolds admitting a non-trivial JSJ decomposition, the case that needs to be understood is that of finite volume hyperbolic 3-manifolds. We focus on this case in the remainder of this section. We first deal with the case of $\mathcal{G}_{\mathfrak{M}}(\pi_1(M))$, where M is a finite volume hyperbolic 3-manifold. In the light of Theorem 4.6, a natural class of hyperbolic 3-manifolds to attempt to establish rigidity for are hyperbolic 3-manifolds that fiber over the circle, since, as we now explain, this can be used to help organize an approach to profinite rigidity of the fundamental groups of hyperbolic 3-manifolds.

Proposition 4.22. *Suppose that for any orientable finite volume hyperbolic 3-manifold M that fibers over the circle we have $\mathcal{G}_{\mathfrak{M}}(\pi_1(M)) = \{\pi_1(M)\}$. Then if N is a finite* *volume hyperbolic 3-manifold and Y a compact 3-manifold with $\pi_1(Y) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(N))$, then Y is commensurable to N.*

Proof. Note that from Corollary 4.19 and Theorem 4.20, Y is a finite volume hyperbolic 3-manifold. By Theorem 4.6, we can pass to finite covers N_f and Y_f of N and Y respectively, that are both fibered, and with $\widehat{\pi_1(N_f)} \cong \widehat{\pi_1(Y_f)}$. By the "rigidity hypothesis" of Proposition 4.22, it follows that $\pi_1(N_f) \cong \pi_1(Y_f)$, and so N and Y share a common finite sheeted cover $N_f \cong Y_f$. □

Thus, it is natural to focus on the case of surface bundles. The following rigidity result is proved in Bridson, Reid, and Wilton [2017] (see also Bridson and Reid [2015b] and Boileau and Friedl [2015] for the the case of the figure-eight knot complement). This is the first family of hyperbolic 3-manifolds that fiber over the circle for which the rigidity required in Proposition 4.22 has been carried to completion. An approach to handle other fibered hyperbolic 3-manifolds is described in Bridson, Reid, and Wilton [2017].

Theorem 4.23. *Let M be a once-punctured torus bundle over the circle (hyperbolic or otherwise). Then $\mathcal{G}_{\mathfrak{M}}(\pi_1(M)) = \{\pi_1(M)\}$.*

Some ideas in the proof: We only discuss the hyperbolic case, and refer the reader to Bridson, Reid, and Wilton [ibid.] for the remaining (simpler) cases. In this case $b_1(M) = 1$. From Theorem 4.20 we can assume that if N is a compact 3-manifold with $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$, then N is a cusped hyperbolic 3-manifold with $b_1(N) = 1$ (recall Remark 3.2). The proof can be broken down into two main steps as follows:

Step 1: Prove that N is fibered with fiber a once-punctured torus.

Step 2: Since M is a once-punctured torus bundle, given Step 1, a simple analysis gives finitely many possibilities for N. Distinguish these finitely many.

We will make no further comment on Step 2 and refer the reader to Bridson, Reid, and Wilton [ibid.]. The proof of Step 1 follows Bridson and Reid [2015b] and we briefly comment on this (a different proof of this is given in Boileau and Friedl [2015]). The main difficulty is in establishing that N is fibered. Once this is done, the fact that the fiber is a once-punctured torus follows routinely.

Note that in Bridson and Reid [2015b] the cases that N is hyperbolic or not hyperbolic were treated separately (since Theorem 4.20 was unavailable at the time of writing). As noted above, using Theorem 4.20 we can now reduce to the case that N is hyperbolic. Regardless of this development, we still need to follow the argument of Bridson and Reid [ibid.] to complete the proof. The key point is that if N is not fibered, then using B. Freedman and M. H. Freedman [1998] we can build a surface subgroup $H < K = \ker\{\pi_1(N) \to \mathbb{Z}\}$ (this homomorphism is unique since $b_1(N) = 1$). By Corollaries 4.8 and 2.9 we deduce that $\widehat{H} \cong \overline{H} < \overline{K} < \widehat{\pi_1(N)}$. Now by uniqueness of the homomorphism $\pi_1(M) \to \mathbb{Z}$, which has kernel a free group F of rank 2, we get $\widehat{H} < \overline{K} \cong \widehat{F}$. However, using cohomological dimension in the context of profinite groups (see Serre [1997]) we get a contradiction: the cohomological dimension of $\widehat{H}$ is 2 and it is 1 for $\widehat{F}$. □

Since Bridson, Reid, and Wilton [2017] was written, the fact that fibering is a profinite invariant has been established by Jaikin-Zapirain [n.d.] without the restriction on $b_1(M)$. The proof of this uses very different methods to those outlined above.

Theorem 4.24 (Jaikin-Zapirain)**.** *Let M be a compact irreducible 3-manifold and let $\Gamma = \pi_1(M)$.*

1. *If $\widehat{\Gamma}$ is isomorphic to the profinite completion of free-by-cyclic group, then M has non-empty boundary consisting of a disjoint union of incompressible tori and Klein bottles, and fibers over the circle with fiber a compact surface with non-empty boundary.*
2. *If $\widehat{\Gamma}$ is isomorphic to the profinite completion of the fundamental group of a closed 3-manifold that fibers over the circle, then M is a surface bundle over the circle with fiber a closed surface.*

One can distill from the cohomological dimension argument used at the end of the proof of Theorem 4.23 the following useful proposition.

Proposition 4.25. *Let Γ be a finitely generated residually finite group that contains a subgroup $H \cong \pi_1(\Sigma_g)$ for some $g \geq 1$ and for which $\overline{H} \cong \widehat{H}$ in $\widehat{\Gamma}$. Then $\Gamma \notin \mathcal{G}(F_n)$ for any $n \geq 2$.*

Remark 4.26. It is worth remarking that $\widehat{F}_n$ contains a subgroup isomorphic to some $\pi_1(\Sigma_g)$ which is dense in $\widehat{F}_n$ (see Breuillard, Gelander, Souto, and Storm [2006]).

4.6 A profinitely rigid Kleinian group. At present it still remains open as to whether there is *any* finite volume hyperbolic 3-manifold $M = \mathbb{H}^3/\Gamma$ with $\mathcal{G}(\Gamma) = \{\Gamma\}$. However in recent work Bridson, McReynolds, Reid, and Spitler [n.d.] if we allow Γ to be a Kleinian group (i.e. a discrete subgroup of $\mathrm{PSL}(2, \mathbb{C})$) containing torsion then this can be done. As far as we can tell, this seems to be first example (indeed we give two) of a group "similar to a free group" that can be proved to be profinitely rigid, and can be viewed as providing the first real evidence towards answering Question 4.3 (and 4.1) in the affirmative. Namely we prove the following theorem in Bridson, McReynolds, Reid, and Spitler [ibid.] (where $\omega^2 + \omega + 1 = 0$).

Theorem 4.27. *The Kleinian groups* $\mathrm{PGL}(2, \mathbb{Z}[!])$ *and* $\mathrm{PSL}(2, \mathbb{Z}[!])$ *are profinitely rigid.*

The case of $\mathrm{PGL}(2, \mathbb{Z}[!])$ follows from that of $\mathrm{PSL}(2, \mathbb{Z}[!])$, and we so we limit ourselves to briefly indicating the strategy of the proof of Theorem 4.27 for $\mathrm{PSL}(2, \mathbb{Z}[!])$.

There are three key steps in the proof which we summarize below.

Theorem 4.28 (Representation Rigidity)**.** *Let $\iota : \Gamma \to \mathrm{PSL}(2, \mathbb{C})$ denote the identity homomorphism, and $c = \bar{\iota}$ the complex conjugate representation. Then if $\rho : \Gamma \to \mathrm{PSL}(2, \mathbb{C})$ is a representation with infinite image, $\rho = \iota$ or c.*

Using Theorem 4.28 we are able to get some control on $\mathrm{PSL}(2,\mathbb{C})$ representations of a finitely generated residually finite group with profinite completion isomorphic to $\widehat{\Gamma}$, and to that end we prove:

Theorem 4.29. *Let Δ be a finitely generated residually finite group with $\widehat{\Delta} \cong \widehat{\Gamma}$. Then Δ admits an epimorphism to a group $L < \Gamma$ which is Zariski dense in* $\mathrm{PSL}(2,\mathbb{C})$.

Finally, we make use of Theorem 4.29, in tandem with an understanding of the topology and deformations of orbifolds $\mathbb{H}^3/G$ for subgroups $G < \Gamma$. Briefly, in the notation of Theorem 4.29, the case of L having infinite index can be ruled out using Teichmüller theory to construct explicit finite quotients of L and hence Δ that cannot be finite quotients of Γ. To rule out the finite index case we make use of information about low-index subgroups of Γ, together with the construction of L, and 3-manifold topology to show that L contains the fundamental group of a once-punctured torus bundle over the circle of index 12. We can then invoke Bridson, Reid, and Wilton [2017] to yield the desired conclusion that $\Delta \cong \Gamma$. □

5 Virtually free groups, Fuchsian groups and Limit groups

We now turn from the world of 3-manifold groups to other classes of groups closely related to free groups; virtually free groups (i.e. contains a free subgroup of finite index), Fuchsian groups which are discrete subgroups of $\mathrm{PSL}(2,\mathbb{R})$ and limit groups which we define below. All three classes of these groups contain the class of free groups amongst them. As already noted even groups that are virtually $\mathbb{Z}$ can fail to be profinitely rigid. In Grunewald and Zalesskii [2011] this is extended to give examples of virtually non-abelian free groups in the same genus, as well as providing cases where they show that certain virtually free groups are the only groups in the genus when restricted to virtually free groups.

5.1 Some restricted genus results. Regarding Fuchsian groups, the following is proved in Bridson, Conder, and Reid [2016].

Theorem 5.1. *Let $\mathcal{L}$ denote the collection of lattices in connected Lie groups and let Γ be a finitely generated Fuchsian group. Then $\mathcal{G}_{\mathcal{L}}(\Gamma) = \{\Gamma\}$.*

Using the profinite invariance of $b_1^{(2)}$ Bridson, Conder, and Reid [ibid.], it turns out that the hard case of Theorem 5.1 is ruling out non-isomorphic Fuchsian having isomorphic profinite completions. The main part of the proof of this step is to rule out "fake torsion" in the profinite completion, and uses the technology of profinite group actions on profinite trees (see Bridson, Conder, and Reid [ibid.] for details).

By a limit group we mean a finitely-generated group Γ that is fully residually free; i.e. a finitely generated group in which every finite subset can be mapped injectively into a free group by a group homomorphism. In connection with Question 4.1, Wilton [2017] recently proved the following:

Theorem 5.2. *Let Γ be a limit group that is not a free group, and let F be a free group. Then $\widehat{\Gamma}$ is not isomorphic to $\widehat{F}$.*

The key point in the proof of Theorem 5.2 (and indeed the main point of Wilton [2017]) is to construct a surface subgroup in a non-free limit group. One can then follow an argument in Bridson, Conder, and Reid [2016] that uses Wilton [2008] (which proves LERF for limit groups) and Proposition 4.25 to complete the proof.

5.2 Profinite genus of free groups and one-ended hyperbolic groups. We close this section with a discussion of possible groups in the genus for free groups. As noted above, Theorem 5.2 uses the existence of surface subgroups to show that non-free limit groups do not lie in the same genus as a free group. The next result from Bridson, Conder, and Reid [2016] takes up this theme, and connects to two well-known open problems about word hyperbolic groups, namely:

(A) Does every 1-ended word-hyperbolic group contain a surface subgroup?

(B) Is every word-hyperbolic group residually finite?

The first question, due to Gromov, was motivated by the case of hyperbolic 3-manifolds, and in this special case the question was settled by Kahn and Markovic [2012]. Indeed, given Kahn and Markovic [ibid.], a natural strengthening of (A) above is to ask:

(A′) Does every 1-ended word-hyperbolic group contain a quasi-convex surface subgroup?

Theorem 5.3. *Suppose that every 1-ended word-hyperbolic group is residually finite and contains a quasi-convex surface subgroup. Then there exists no 1-ended word-hyperbolic group Γ and free group F such that $\widehat{\Gamma} \cong \widehat{F}$.*

Proof. Assume the contrary, and let Γ be a counter-example, with $\widehat{\Gamma} \cong \widehat{F}$ for some free group F. Let H be a quasi-convex surface subgroup of Γ. Note that the finite-index subgroups of H are also quasi-convex in Γ. Under the assumption that all 1-ended hyperbolic groups are residually finite, it is proved in Agol, Groves, and Manning [2009] that H and all its subgroups of finite index must be separable in Γ. Hence by Lemma 2.8, the natural map $\widehat{H} \to \overline{H} < \widehat{\Gamma} \cong \widehat{F}$ is an isomorphism, and can use Proposition 4.25 to complete the proof. □

Corollary 5.4. *Suppose that there exists a 1-ended word hyperbolic group Γ with $\widehat{\Gamma} \cong \widehat{F}$ for some free group F. Then either there exists a word-hyperbolic group that is not residually finite, or there exists a word-hyperbolic group that does not contain a quasi-convex surface subgroup.*

6 Profinite rigidity and flexibility in other settings

Although our attention has been on groups arising from low-dimensional geometry and topology we think it worthwhile to include a (far from complete) survey of profinite rigidity and flexibility for other classes of finitely generated or finitely presented groups.

6.1 Nilpotent and polycyclic groups. As is already evident from Baumslag's examples of meta-cyclic groups in Section 3, the case of nilpotent groups already shows some degree of subtlety. The case of nilpotent groups more generally is well understood due to work of Pickel [1971]. We will not discuss this in any detail, other than to say that, in Pickel [ibid.] it is shown that for a finitely generated nilpotent group Γ, $\mathcal{G}(\Gamma)$ consists of a finite number of isomorphism classes of nilpotent groups, and moreover, examples where the genus can be made arbitrarily large are known (see for example Segal [1983] Chapter 11). Examples of profinitely rigid nilpotent groups of class 2 are constructed in Grunewald and Scharlau [1979].

Similar results are also known for polycyclic groups and we refer the reader to Grunewald and Segal [1978] and Segal [1983]. Note that in the case of nilpotent groups it is straightforward to prove that any finitely generated residually finite group in the same genus as a nilpotent group is nilpotent. The same holds for polycyclic groups (see Sabbagh and Wilson [1991]), but this is a good deal harder.

These results should be compared with the examples of the meta-abelian groups (which are solvable) of Pickel given in Remark 3.4 where the genus is infinite.

6.2 Lattices in semi-simple Lie groups. Let Γ be a lattice in a semi-simple Lie group, for example, in what follows we shall take $\Gamma = \mathrm{SL(n, R_k)}$ where R_k denotes the ring of integers in a number field k. A natural, obvious class of finite quotients of Γ, are those of the form $\mathrm{SL(n, R_k/I)}$ where $I \subset R_k$ is an ideal. Let $\Gamma(I)$ denote the kernel of the reduction homomorphism $\Gamma \to \mathrm{SL(n, R_k/I)}$. By Strong Approximation for $\mathrm{SL_n}$ (see Platonov and Rapinchuk [1994] Chapter 7.4 for example) these reduction homomorphisms are surjective for all I. A *congruence subgroup* of Γ is any subgroup $\Delta < \Gamma$ such that $\Gamma(I) < \Delta$ for some I. A group Γ is said to have the *Congruence Subgroup Property* (abbreviated to CSP) if every subgroup of finite index is a congruence subgroup.

Thus, if Γ has CSP, then $\mathcal{C}(\Gamma)$ is known precisely, and in effect, to determine $\mathcal{C}(\Gamma)$ is reduced to number theory. Expanding on this, since R_k is a Dedekind domain, any ideal I factorizes into powers of prime ideals. If $I = \prod \mathcal{P}_i^{a_i}$, then it is known that $\mathrm{SL(n, R_k/I)} = \prod \mathrm{SL(n, R_k/\mathcal{P}_i^{a_i})}$. Thus the finite groups that arise as quotients of $\mathrm{SL(n, R_k)}$ are determined by those of the form $\mathrm{SL(n, R_k/\mathcal{P}_i^{a_i})}$. Hence we are reduced to understanding how a rational prime p behaves in the extension $k/\mathbb{Q}$. This idea, coupled with the work of Serre [1970] which has shed considerable light on when Γ has CSP allows construction of non-isomorphic lattices in the same genus. We refer the reader to Aka [2012b], Aka [2012a] and Reid [2015] for further details.

6.3 Grothendieck's Problem. A particular case of when discrete groups groups have isomorphic profinite completions is the following (which goes back to Grothendieck [1970]).

Let Γ be a residually finite group and let $u : P \hookrightarrow \Gamma$ be the inclusion of a subgroup P. Then (Γ, P) is called a *Grothendieck Pair* if the induced homomorphism $\widehat{u} : \widehat{P} \to \widehat{\Gamma}$ is an isomorphism but u is not.

We say that Γ is *Grothendieck Rigid* if no proper finitely generated subgroup $u : P \to \Gamma$ gives a Grothendieck Pair.

Grothendieck [1970] asked about the existence of Grothendieck Pairs of finitely presented groups and the first such pairs were constructed by Bridson and Grunewald [2004]. The analogous problem for finitely generated groups had been settled earlier by Platonov and Tavgen [1990] (see also Bass and Lubotzky [2000]). Using different methods, Pyber [2004] gave a construction of continuously many finitely generated groups Γ_α with subgroups H_α for which $(\Gamma_\alpha, H_\alpha)$ are Grothendieck Pairs.

The constructions of Platonov and Tavgen [1990] and Bridson and Grunewald [2004] rely on versions of the following result (see also Bridson [2010]). We remind the reader that the *fibre product* $P < \Gamma \times \Gamma$ associated to an epimorphism of groups $p : \Gamma \to Q$ is the subgroup $P = \{(x, y) \mid p(x) = p(y)\}$.

Proposition 6.1. *Let $1 \to N \to \Gamma \to Q \to 1$ be a short exact sequence of groups with Γ finitely generated and let P be the associated fibre product. Suppose that $Q \neq 1$ is finitely presented, has no proper subgroups of finite index, and $H_2(Q, \mathbb{Z}) = 0$. Then*

1. *$(\Gamma \times \Gamma, P)$ is a Grothendieck Pair;*
2. *if N is finitely generated then (Γ, N) is a Grothendieck Pair.*

More recently in Bridson [2016], examples of Grothendieck Pairs were constructed so as to provide the first examples of finitely-presented residually finite groups Γ that contain an infinite sequence of non-isomorphic finitely presented subgroups P_n so that (Γ, P_n) are Grothendieck Pairs. In particular, this provides examples of finitely presented groups Γ for which $\mathcal{G}(\Gamma)$ is infinite. These examples are non-solvable in contrast to those of Pickel in Remark 3.4

Note that if a H is a subgroup of a group Γ and Γ is H-separable then it is easy to see that (Γ, H) cannot be a Grothendieck Pair (since H is not dense in the profinite topology). This was noticed in Platonov and Tavgen [1990] to observe that free groups and Fuchsian groups were Grothendieck Rigid. For 3-manifolds Grothendieck Rigidity was shown in Long and Reid [2011] for the fundamental groups of closed geometric 3-manifolds and finite volume hyperbolic 3-manifolds without appealing to LERF in the hyperbolic case. In Cavendish [2012] and Reid [2015] this was extended to the fundamental groups of all closed irreducible 3-manifolds (as a consequence of Theorem 4.11). This program has been completed by Boileau and Friedl [2017] who proved:

Theorem 6.2. *The fundamental group of any compact, connected, irreducible, orientable 3-manifold with empty or toroidal boundary is Grothendieck Rigid.*

7 Final remarks and further questions

As should be clear from this article, the questions posed in Section 4 remain stubbornly open, and even questions about the nature of $\mathcal{G}_{\mathfrak{M}}(\pi_1(M))$ for M a finite volume hyperbolic 3-manifold seem hard to resolve. Never the less, these open problems can be used as platforms for other, perhaps more approachable problems. We discuss a few, other problems for other classes of groups can be found in Reid [2015].

Question 7.1. Let Γ denote the fundamental group of the figure-eight knot complement. It is well-known that Γ has index 12 in the group $\mathrm{PSL}(2, \mathbb{Z}[!])$ of Theorem 4.27. Is Γ is profinitely rigid?

As noted in Section 4.5, it was shown in Bridson and Reid [2015b] and Boileau and Friedl [2015] that $\mathcal{G}_{\mathfrak{M}}(\Gamma) = \{\Gamma\}$.

Question 7.2. Let M_W denote the Weeks manifold. This is the smallest volume hyperbolic 3-manifold Gabai, Meyerhoff, and Milley [2009]. Is $\mathcal{G}_{\mathfrak{M}}(\pi_1(M_W) = \{\pi_1(M_W)\}$?

Indeed, one might wonder whether the techniques of Bridson, McReynolds, Reid, and Spitler [n.d.] (as described in Theorem 4.27) can be brought to bear in this example since $\pi_1(M_W)$ exhibits a certain amount of representation rigidity.

Question 7.3. In Section 4.5 it was pointed out that recently Jaikin-Zapirain [n.d.] showed that being fibered is a profinite invariant. Given this, a natural question is:

Is the Thurston norm ball a profinite invariant? That is to say, if M is a closed hyperbolic 3-manifold and N a closed hyperbolic 3-manifold with $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$ are the Thurston norm balls isomorphic?

Some progress on this is given in Boileau and Friedl [2015] under an additional condition on the isomorphsim between profinite completions. However, it seems unlikely that this condition will hold in general.

Question 7.4. Is the volume a profinite invariant? That is to say, if M is a finite volume hyperbolic 3-manifold and N a finite volume hyperbolic 3-manifold with $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$ does $\mathrm{vol}(M) = \mathrm{vol}(N)$?

It follows from well-known properties of the set of volumes of hyperbolic 3-manifolds Thurston [1979] that if Question 7.4 has a positive answer then $|\mathcal{G}_{\mathfrak{M}}(\pi_1(M))|$ is finite.

There does appear to be some conjectural evidence to support a positive answer. Briefly, it is conjectured (roughly) that if $\{\Gamma_n\}$ is a cofinal sequence of subgroups of finite index in $\pi_1(M)$ (as above), then:

$$\frac{\log|\mathrm{Tor}(H_1(\Gamma_n, \mathbb{Z}))|}{[\pi_1(M) : \Gamma_n]} \to \frac{1}{6\pi}\mathrm{vol}(M) \text{ as } n \to \infty.$$

Note that $\mathrm{Tor}(H_1(\Gamma_n, \mathbb{Z}))$ is visible in the profinite completions $\widehat{\Gamma}_n$ and so if the above conjecture is true, this would allow one to deduce $\pi_1(N) \in \mathcal{G}_{\mathfrak{M}}(\pi_1(M))$ implies $\mathrm{vol}(M) = \mathrm{vol}(N)$.

References

Ian Agol (2008). "Criteria for virtual fibering". *J. Topol.* 1.2, pp. 269–284. MR: 2399130 (cit. on p. 1217).

– (2013). "The virtual Haken conjecture". *Doc. Math.* 18. With an appendix by Agol, Daniel Groves, and Jason Manning, pp. 1045–1087. MR: 3104553 (cit. on pp. 1211, 1217–1219).

Ian Agol, Daniel Groves, and Jason Fox Manning (2009). "Residual finiteness, QCERF and fillings of hyperbolic groups". *Geom. Topol.* 13.2, pp. 1043–1073. MR: 2470970 (cit. on p. 1226).

Menny Aka (2012a). "Arithmetic groups with isomorphic finite quotients". *J. Algebra* 352, pp. 322–340. MR: 2862189 (cit. on p. 1227).

– (2012b). "Profinite completions and Kazhdan's property (T)". *Groups Geom. Dyn.* 6.2, pp. 221–229. MR: 2914858 (cit. on p. 1227).

Matthias Aschenbrenner, Stefan Friedl, and Henry Wilton (2015). *3-manifold groups*. EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, pp. xiv+215. MR: 3444187 (cit. on p. 1218).

Hyman Bass and Alexander Lubotzky (2000). "Nonarithmetic superrigid groups: counterexamples to Platonov's conjecture". *Ann. of Math. (2)* 151.3, pp. 1151–1173. MR: 1779566 (cit. on p. 1228).

Gilbert Baumslag (1974). "Residually finite groups with the same finite images". *Compositio Math.* 29, pp. 249–252. MR: 0357615 (cit. on pp. 1215, 1216).

Michel Boileau and Stefan Friedl (May 2015). "The profinite completion of 3-manifold groups, fiberedness and the Thurston norm". arXiv: 1505.07799 (cit. on pp. 1223, 1229).

– (Oct. 2017). "Grothendieck rigidity of 3-manifold groups". arXiv: 1710.02746 (cit. on p. 1228).

Francis Bonahon (2002). "Geometric structures on 3-manifolds". In: *Handbook of geometric topology*. North-Holland, Amsterdam, pp. 93–164. MR: 1886669 (cit. on p. 1217).

Emmanuel Breuillard, Tsachik Gelander, Juan Souto, and Peter Storm (2006). "Dense embeddings of surface groups". *Geom. Topol.* 10, pp. 1373–1389. MR: 2255501 (cit. on p. 1224).

Martin R. Bridson (2010). "The Schur multiplier, profinite completions and decidability". *Bull. Lond. Math. Soc.* 42.3, pp. 412–416. MR: 2651935 (cit. on p. 1228).

– (2016). "The strong profinite genus of a finitely presented group can be infinite". *J. Eur. Math. Soc. (JEMS)* 18.9, pp. 1909–1918. MR: 3531665 (cit. on p. 1228).

Martin R. Bridson, M. D. E. Conder, and Alan W. Reid (2016). "Determining Fuchsian groups by their finite quotients". *Israel J. Math.* 214.1, pp. 1–41. MR: 3540604 (cit. on pp. 1221, 1225, 1226).

Martin R. Bridson and Fritz J. Grunewald (2004). "Grothendieck's problems concerning profinite completions and representations of groups". *Ann. of Math. (2)* 160.1, pp. 359–373. MR: 2119723 (cit. on p. 1228).

Martin R. Bridson, D. B. McReynolds, Alan W. Reid, and R. Spitler (n.d.). "Profinite rigidity of $\mathbf{PGL}(2, \mathbb{Z}[\omega])$ and $\mathbf{PSL}(2, \mathbb{Z}[\omega])$". In preparation (cit. on pp. 1224, 1229).

Martin R. Bridson and Alan W. Reid (2015a). "Nilpotent completions of groups, Grothendieck pairs, and four problems of Baumslag". *Int. Math. Res. Not. IMRN* 8, pp. 2111–2140. MR: 3344664 (cit. on p. 1222).

– (May 2015b). "Profinite rigidity, fibering, and the figure-eight knot". arXiv: 1505.07886 (cit. on pp. 1223, 1229).

Martin R. Bridson, Alan W. Reid, and Henry Wilton (2017). "Profinite rigidity and surface bundles over the circle". *Bull. Lond. Math. Soc.* 49.5, pp. 831–841. MR: 3742450 (cit. on pp. 1223–1225).

Will Cavendish (2012). *Finite-Sheeted Covering Spaces and Solenoids over 3-manifolds*. Thesis (Ph.D.)–Princeton University. ProQuest LLC, Ann Arbor, MI, p. 96. MR: 3078440 (cit. on pp. 1219, 1228).

John D. Dixon, Edward W. Formanek, John C. Poland, and Luis Ribes (1982). "Profinite completions and isomorphic finite quotients". *J. Pure Appl. Algebra* 23.3, pp. 227–231. MR: 644274 (cit. on p. 1213).

Benedict Freedman and Michael H. Freedman (1998). "Kneser-Haken finiteness for bounded 3-manifolds locally free groups, and cyclic covers". *Topology* 37.1, pp. 133–147. MR: 1480882 (cit. on p. 1223).

Louis Funar (2013). "Torus bundles not distinguished by TQFT invariants". *Geom. Topol.* 17.4. With an appendix by Funar and Andrei Rapinchuk, pp. 2289–2344. MR: 3109869 (cit. on pp. 1220, 1221).

David Gabai (1986). "On 3-manifolds finitely covered by surface bundles". In: *Low-dimensional topology and Kleinian groups (Coventry/Durham, 1984)*. Vol. 112. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 145–155. MR: 903863 (cit. on p. 1217).

David Gabai, Robert Meyerhoff, and Peter Milley (2009). "Minimum volume cusped hyperbolic three-manifolds". *J. Amer. Math. Soc.* 22.4, pp. 1157–1215. MR: 2525782 (cit. on p. 1229).

Alexander Grothendieck (1970). "Représentations linéaires et compactification profinie des groupes discrets". *Manuscripta Math.* 2, pp. 375–396. MR: 0262386 (cit. on pp. 1227, 1228).

Daniel Groves and Jason Fox Manning (Aug. 2017). "Quasiconvexity and Dehn filling". arXiv: 1708.07968 (cit. on p. 1217).

Fritz J. Grunewald, A. Jaikin-Zapirain, and Pavel A. Zalesskii (2008). "Cohomological goodness and the profinite completion of Bianchi groups". *Duke Math. J.* 144.1, pp. 53–72. MR: 2429321 (cit. on pp. 1218, 1219).

Fritz J. Grunewald and Rudolf Scharlau (1979). "A note on finitely generated torsion-free nilpotent groups of class 2". *J. Algebra* 58.1, pp. 162–175. MR: 535851 (cit. on p. 1227).

Fritz J. Grunewald and Daniel Segal (1978). "On polycyclic groups with isomorphic finite quotients". *Math. Proc. Cambridge Philos. Soc.* 84.2, pp. 235–246. MR: 0491962 (cit. on p. 1227).

Fritz J. Grunewald and Pavel A. Zalesskii (2011). "Genus for groups". *J. Algebra* 326, pp. 130–168. MR: 2746056 (cit. on p. 1225).

John Hempel (Sept. 2014). "Some 3-manifold groups with the same finite quotients". arXiv: 1409.3509 (cit. on p. 1220).

R. Hirshon (1977). "Some cancellation theorems with applications to nilpotent groups". *J. Austral. Math. Soc. Ser. A* 23.2, pp. 147–165. MR: 0447410 (cit. on p. 1216).

A. Jaikin-Zapirain (n.d.). "Recogniton of fibering in compact 3-manifolds". In preparation (cit. on pp. 1224, 1229).

Jeremy Kahn and Vladimir Markovic (2012). "Immersing almost geodesic surfaces in a closed hyperbolic three manifold". *Ann. of Math. (2)* 175.3, pp. 1127–1190. MR: 2912704 (cit. on p. 1226).

Sławomir Kwasik and Witold Rosicki (2004). "On stability of 3-manifolds". *Fund. Math.* 182.2, pp. 169–180. MR: 2100626 (cit. on p. 1220).

Darren D. Long and Alan W. Reid (2011). "Grothendieck's problem for 3-manifold groups". *Groups Geom. Dyn.* 5.2, pp. 479–499. MR: 2782181 (cit. on p. 1228).

John Lott and Wolfgang Lück (1995). "L^2-topological invariants of 3-manifolds". *Invent. Math.* 120.1, pp. 15–60. MR: 1323981 (cit. on p. 1221).

Alexander Lubotzky (1993). "Torsion in profinite completions of torsion-free groups". *Quart. J. Math. Oxford Ser. (2)* 44.175, pp. 327–332. MR: 1240475 (cit. on p. 1219).

W. Lück (1994). "Approximating L^2-invariants by their finite-dimensional analogues". *Geom. Funct. Anal.* 4.4, pp. 455–481. MR: 1280122 (cit. on p. 1221).

John Morgan and Gang Tian (2014). *The geometrization conjecture*. Vol. 5. Clay Mathematics Monographs. American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, pp. x+291. MR: 3186136 (cit. on p. 1217).

Walter D. Neumann (1997). "Commensurability and virtual fibration for graph manifolds". *Topology* 36.2, pp. 355–378. MR: 1415593 (cit. on p. 1218).

Nikolay Nikolov and Dan Segal (2007). "On finitely generated profinite groups. I. Strong completeness and uniform bounds". *Ann. of Math. (2)* 165.1, pp. 171–238. MR: 2276769 (cit. on p. 1213).

G. A. Noskov, V. N. Remeslennikov, and V. A. Romankov (1979). "Infinite groups". In: *Algebra. Topology. Geometry, Vol. 17 (Russian)*. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow, pp. 65–157, 308. MR: 584569 (cit. on p. 1216).

Grisha Perelman (Nov. 2002). "The entropy formula for the Ricci flow and its geometric applications". arXiv: math/0211159.

– (Mar. 2003). "Ricci flow with surgery on three-manifolds". arXiv: math/0303109.

P. F. Pickel (1971). "Finitely generated nilpotent groups with isomorphic finite quotients". *Trans. Amer. Math. Soc.* 160, pp. 327–341. MR: 0291287 (cit. on p. 1227).

– (1974). "Metabelian groups with the same finite quotients". *Bull. Austral. Math. Soc.* 11, pp. 115–120. MR: 0364455 (cit. on p. 1216).

Vladimir P. Platonov and Andrei Rapinchuk (1994). *Algebraic groups and number theory*. Vol. 139. Pure and Applied Mathematics. Translated from the 1991 Russian original by Rachel Rowen. Academic Press, Inc., Boston, MA, pp. xii+614. MR: 1278263 (cit. on p. 1227).

Vladimir P. Platonov and O. I. Tavgen (1990). "Grothendieck's problem on profinite completions and representations of groups". *K-Theory* 4.1, pp. 89–101. MR: 1076526 (cit. on p. 1228).

Piotr Przytycki and Daniel T. Wise (2018). "Mixed 3-manifolds are virtually special". *J. Amer. Math. Soc.* 31.2, pp. 319–347. MR: 3758147 (cit. on p. 1218).

László Pyber (2004). "Groups of intermediate subgroup growth and a problem of Grothendieck". *Duke Math. J.* 121.1, pp. 169–188. MR: 2031168 (cit. on p. 1228).

Alan W. Reid (2015). "Profinite properties of discrete groups". In: *Groups St Andrews 2013*. Vol. 422. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 73–104. MR: 3445488 (cit. on pp. 1219, 1227, 1228).

Luis Ribes and Pavel A. Zalesskii (2000). *Profinite groups*. Vol. 40. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, Berlin, pp. xiv+435. MR: 1775104 (cit. on pp. 1212, 1213).

Gabriel Sabbagh and John S. Wilson (1991). "Polycyclic groups, finite images, and elementary equivalence". *Arch. Math. (Basel)* 57.3, pp. 221–227. MR: 1119894 (cit. on p. 1227).

Peter Scott (1978). "Subgroups of surface groups are almost geometric". *J. London Math. Soc. (2)* 17.3. See correction in Scott [1985], pp. 555–565. MR: 0494062 (cit. on pp. 1218, 1233).

– (1985). "Correction to: Subgroups of surface groups are almost geometric". *J. London Math. Soc. (2)* 32.2. Correction to Scott [1978], pp. 217–220. MR: 811778 (cit. on p. 1233).

Daniel Segal (1983). *Polycyclic groups*. Vol. 82. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, pp. xiv+289. MR: 713786 (cit. on p. 1227).

Jean-Pierre Serre (1970). "Le problème des groupes de congruence pour SL_2". *Ann. of Math. (2)* 92, pp. 489–527. MR: 0272790 (cit. on p. 1227).

– (1997). *Galois cohomology*. Translated from the French by Patrick Ion and revised by the author. Springer-Verlag, Berlin, pp. x+210. MR: 1466966 (cit. on pp. 1218, 1224).

Hongbin Sun (Aug. 2016). "NonLERFness of arithmetic hyperbolic manifold groups". arXiv: 1608.04816 (cit. on p. 1218).

William P. Thurston (1979). "The Geometry and Topology of 3-Manifolds". Princeton University mimeographed notes (cit. on p. 1229).

– (1997). *Three-dimensional geometry and topology. Vol. 1*. Vol. 35. Princeton Mathematical Series. Edited by Silvio Levy. Princeton University Press, Princeton, NJ, pp. x+311. MR: 1435975 (cit. on p. 1217).

Gareth Wilkes (May 2016). "Profinite rigidity of graph manifolds and JSJ decompositions of 3-manifolds". arXiv: 1605.08244 (cit. on p. 1221).

– (2017). "Profinite rigidity for Seifert fibre spaces". *Geom. Dedicata* 188, pp. 141–163. MR: 3639628 (cit. on p. 1220).

Henry Wilton (2008). "Hall's theorem for limit groups". *Geom. Funct. Anal.* 18.1, pp. 271–303. MR: 2399104 (cit. on p. 1226).

– (Jan. 2017). "Essential surfaces in graph pairs". arXiv: 1701.02505 (cit. on pp. 1225, 1226).

Henry Wilton and Pavel A. Zalesskii (2017a). "Distinguishing geometries using finite quotients". *Geom. Topol.* 21.1, pp. 345–384. MR: 3608716 (cit. on pp. 1220–1222).

– (2017b). "Pro-*p* subgroups of profinite completions of 3-manifold groups". *J. Lond. Math. Soc. (2)* 96.2, pp. 293–308. arXiv: 1703 . 03701. MR: 3708951 (cit. on pp. 1221, 1222).

Daniel T. Wise (2009). "Research announcement: the structure of groups with a quasiconvex hierarchy". *Electron. Res. Announc. Math. Sci.* 16, pp. 44–55. MR: 2558631 (cit. on pp. 1211, 1217–1219, 1222).

Received 2017-11-29.

Alan W. Reid
Department of Mathematics
Rice University
Houston, TX 77005, USA
alan.reid@rice.edu

video: slides:

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1235–1258)

THE EVENNESS CONJECTURE IN EQUIVARIANT UNITARY BORDISM

Bernardo Uribe

Abstract

The evenness conjecture for the equivariant unitary bordism groups states that these bordism groups are free modules over the unitary bordism ring in even dimensional generators. In this paper we review the cases on which the conjecture is known to hold and we highlight the properties that permit to prove the conjecture in these cases.

Introduction

The G-equivariant unitary bordism groups for G a compact Lie group are the bordism groups of G-equivariant tangentially stable almost complex manifolds, also known as G-equivariant unitary manifolds. These are closed G-manifolds M for which a stable tangent bundle $TM \oplus \underline{\mathbb{R}}^k$, where $\underline{\mathbb{R}}^k$ denotes the trivial bundle $\mathbb{R}^k \times M$ with trivial G-action, can be endowed with the structure of a G-equivariant complex bundle. Two tangentially stable almost complex G-structures are identified if after stabilization with further G-trivial $\underline{\mathbb{C}}$ summands the structures become G-homotopic through complex G-structures. Being unitary is inherited to the fixed points sets. Whenever H is a closed subgroup of G the fixed points M^H are also tangentially stable almost complex, and moreover a N_H-tubular neighborhood around M^H in M possesses a complex N_H structure May [1996, §XVIII, Prop. 3.2].

For a cofibration of G-spaces $Y \to X$, the geometric G-equivariant unitary bordism groups $\Omega_n^G(X, Y)$ are the G-bordism classes of G-equivariant n-dimensional manifolds with map $(M^n, \partial M^n) \to (X, Y)$.

The author acknowledges and thanks the financial support provided by the Max Planck Institute for Mathematics, by COLCIENCIAS through the Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación and by the Alexander Von Humboldt Foundation.
MSC2010: primary 55N22; secondary 57R77, 57R85.
Keywords: Equivariant unitary bordism, equivariant vector bundle.

The G-equivariant unitary bordism groups of a point Ω_*^G become a ring under the cartesian product of manifolds with the diagonal G-action, and therefore a module of the unitary bordism ring Ω_* whenever we consider a unitary manifold as a trivial G-manifold. Milnor [1960] and Novikov [1960], using Thom's remarkable calculation of the unoriented bordism groups Thom [1954], showed that the unitary bordism ring is a polinomial ring $\Omega_* = \mathbb{Z}[x_{2i} : i \geq 1]$ with one generator in each even degree. In this work we will be interested in the Ω_*-module structure of the equivariant unitary bordism groups Ω_*^G.

Explicit calculations carried out by Landweber [1972] in the cyclic case and by Stong [1970] in the abelian p-group case permitted them to conclude that in these cases the equivariant unitary bordism group Ω_*^G is a free Ω_*-module in even dimensional generators. Ossa [1972] generalized this result to any finite abelian group and Löffler [1974] and Comezaña in May [1996, §XXVIII, Thm. 5.1] showed that this also holds whenever G is a compact abelian Lie group. Explicit calculations done for the Diehadral groups D_{2p} with p prime by Ángel, Gómez, and B. Uribe [n.d.] for groups of order pq with p and q different primes by Lazarov [1972] and for metacyclic groups by Rowlett [1980] show that for these groups this phenomenon also occurs. We believe that this property should hold in the G-equivariant unitary bordism groups for any compact Lie group G, in the same way that the coefficients for G-equivariant K-theory are trivial in odd degrees and a free module over the integers in even degrees.

The theme of this work is the

Evenness conjecture for the equivariant unitary bordism groups

which states that the G-equivariant unitary bordism group is a free Ω_*-modules in even dimensional generators whenever G is a compact Lie group. Rowlett explicitly mentions this conjecture in his work of 1980 Rowlett [ibid.] and later Comezaña in his work of 1996 May [1996, §XXVIII.5]. We also believe that this conjecture holds in general and we do hope that this paper will help spreading it to the mathematical community for its eventual proof.

In this work we survey the original proofs of the known cases of the evenness conjecture for finite groups. We start in Section 1 with the definition of the equivariant unitary bordism groups for pairs of families and the long exact sequence associated to them. In Section 2 we review the decomposition of equivariant complex vector bundles restricted to fixed point sets done by Ángel, Gómez, and B. Uribe [n.d.] and how this decomposition allows to write the equivariant unitary bordism groups of adjacent pair of families as the bordism groups of equivariant classifying spaces. In Section 3 we review the proofs of the evenness conjecture for done by Landweber [1972] for cyclic groups, by Stong [1970] for abelian p-groups and by Ossa [1972] for general finite abelian groups. In Section 4 we review the proof of the evenness conjecture for metacyclic groups done by Rowlett [1980] and we finish with in Section 5 with some conclusions.

1 Equivariant unitary bordism for families of subgroups

To study the equivariant bordism groups Conner and Floyd introduced the study of bordism groups of manifolds with prescribed isotropy groups Conner and Floyd [1966, §5]. A family of subgroups $\mathfrak{F}$ of G is a set of subgroups of G which is closed under taking subgroups and under conjugation. The classifying space for the family $E\mathfrak{F}$ is a G-space which is terminal in the category of $\mathfrak{F}$-numerable G-spaces tom Dieck [1987, §1, Thm 6.6] and characterized by the following properties on fixed point sets: $E\mathfrak{F}^H \simeq *$ if $H \in \mathfrak{F}$ and $E\mathfrak{F}^H = \varnothing$ if $H \notin \mathfrak{F}$. This classifying space may be constructed in such a way that whenever $\mathfrak{F}' \subset \mathfrak{F}$, the induced map $E\mathfrak{F}' \to E\mathfrak{F}$ is a G-cofibration.

The equivariant unitary bordism groups for pairs of families may be defined as follows

$$\Omega_*^G[\mathfrak{F}, \mathfrak{F}'](X, A) := \Omega_*^G(X \times E\mathfrak{F}, X \times E\mathfrak{F}' \cup A \times E\mathfrak{F}'),$$

see tom Dieck [1972, p. 310], or alternatively they may be defined in a geometric way is in Stong [1970, §2].

A $(\mathfrak{F}, \mathfrak{F}')$ *free geometric unitary bordism element* of (X, A) is an equivalence class of 4-tuples (M, M_0, M_1, f), where:

- M is an n-dimensional G-manifold endowed with tangentially stable almost G structure which is moreover $\mathfrak{F}$-free, i.e. such that all isotropy groups $G_m = \{g \in G | gm = m\}$ for $m \in M$ belong to $\mathfrak{F}$, and such that $f : M \to X$ is G-equivariant; and

- M_0, M_1 are compact submanifolds of the boundary of M, with $\partial M = M_0 \cup M_1$, $M_0 \cap M_1 = \partial M_0 = \partial M_1$ having tangentially stable almost complex structure induced from M, both G-invariant, such that $f(M_1) \subset A$ and M_0 if $\mathfrak{F}'$-free, i.e. all isotropy groups of M_0 belong to $\mathfrak{F}'$.

Two four tuples (M, M_0, M_1, f) and (M', M_0', M_1', f) are equivalent if there is a 5-tuple (V, V^+, V_0, V_1, F) where

- V is a $\mathfrak{F}$-free manifold and $F : V \to X$ is a G-equivariant map;

- The boundary of V is the union of M, M' and V^+ with $M \cap V^+ = \partial M$, $M' \cap V^+ = \partial M'$, $M \cap M' = \varnothing$, $V^+ \cap (M \cup M') = \partial V^+$, with V inducing the tangentially stable almost complex G-structure on M and the opposite one on M'; V^+ is G-invariant and F restricts to f in M and to f' on M'; and

- V^+ is the union of the G-invariant submanifolds V_0, V_1 with intersection a submanifold V^- in their boundaries, such that $\partial V_i = M_i \cup V^- \cup M_i'$, $M_i \cap V^- = \partial M_i$, $M_i' \cap V^- = \partial M_i'$ with V_0 is $\mathfrak{F}'$-free and $F(V_1) \subset A$.

Definition 1.1. The set of equivalence classes of $(\mathfrak{F}, \mathfrak{F}')$-free geometric unitary bordism elements of (X, A), consisting of clases (M, M_0, M_1, f) where the dimension of M is n, and under the operation of disjoint union, forms an abelian group denoted by

$$\Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A).$$

Call these groups the geometric G-equivariant unitary bordism groups of the pair (X, A) restricted to the pair of families $\mathfrak{F}' \subset \mathfrak{F}$.

Note that if N is a tangentially stable almost complex closed manifold, we can define $N \cdot (M, M_0, M_1, f) = (N \times M, N \times M_0, N \times M_1, f \circ \pi_M)$ thus making $\Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A)$ a module over the unitary bordism ring Ω_*.

The covariant functor $\Omega_*^G\{\mathfrak{F}, \mathfrak{F}'\}$ defines a G-equivariant homology theory Stong [1970, Prop. 2.1], the boundary map on A

$$\begin{aligned} \delta : \Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A) &\to \Omega_{n-1}^G\{\mathfrak{F}, \mathfrak{F}'\}(A, \varnothing) \\ (M, M_0, M_1, f) &\mapsto (M_1, \partial M_1, \varnothing, f|_{M_1}) \end{aligned}$$

defines the long exact sequence in homology for pairs

$$\cdots \Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A) \xrightarrow{\delta} \Omega_{n-1}^G\{\mathfrak{F}, \mathfrak{F}'\}(A, \varnothing) \to \Omega_{n-1}^G\{\mathfrak{F}, \mathfrak{F}'\}(X, \varnothing) \to \cdots,$$

and for families $\mathfrak{F}'' \subset \mathfrak{F}' \subset \mathfrak{F}$, choosing the boundary which is $\mathfrak{F}'$-free

$$\begin{aligned} \partial : \Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A) &\to \Omega_{n-1}^G\{\mathfrak{F}', \mathfrak{F}''\}(X, A) \\ (M, M_0, M_1, f) &\mapsto (M_0, \varnothing, \partial M_0, f|_{M_0}) \end{aligned}$$

one obtains by Stong [ibid., Prop. 2.2] the long exact sequence in homology for families

$$\cdots \Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A) \xrightarrow{\partial} \Omega_{n-1}^G\{\mathfrak{F}', \mathfrak{F}''\}(X, A) \to \Omega_{n-1}^G\{\mathfrak{F}, \mathfrak{F}''\}(X, A) \to \cdots.$$

The bordism condition restricted to the non-relative case $\Omega_*^G\{\mathfrak{F}, \mathfrak{F}'\}(X)$ can be read as the set bordism classes of maps $f : M \to X$ such that M is $\mathfrak{F}$-free and ∂M is $\mathfrak{F}'$-free with M endowed with a tangentially stable almost complex G-structure. Two become equivalent if there exists a G-manifold $F : V \to X$ which is $\mathfrak{F}$-free such that $\partial V = M \cup M' \cup V^+$ and $M \cap V^+ = \partial M$, $M' \cap V^+ = \partial M'$, $M \cap M' = \varnothing$, $V^+ \cap (M \cup M') = \partial V^+$, with the property that F restricts to f on M and to f' on M' and with V^+ $\mathfrak{F}'$-free.

In tom Dieck [1972, Satz 3] it is shown that the canonical map that one can define

$$\mu : \Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X, A) \to \Omega_n^G[\mathfrak{F}, \mathfrak{F}'](X, A) \tag{1.2}$$

becomes a natural isomorphism of homology theories.

A key fact about the $(\mathfrak{F}, \mathfrak{F}')$-free geometric unitary bordism elements of X is the following result proven in Conner and Floyd [1966, Lemma 5.2]. Whenever $(M^n, \partial M^n, f)$ is a $(\mathfrak{F}, \mathfrak{F}')$-free geometric unitary bordism element of X and W^n a compact manifold with boundary regularly embedded in the interior of M^n and invariant under the G-action, such that $G_m \in \mathfrak{F}'$ for all $m \in M^n \backslash W^n$, then $[M^n, \partial M^n, f] = [W^n, \partial W_n, f|_{W^n}]$ in $\Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X)$.

Whenever the pair of families $\mathfrak{F}' \subset \mathfrak{F}$ differ by a fixed group A, i.e. $\mathfrak{F} \backslash \mathfrak{F}' = (A)$ with (A) the set of subgroups of G conjugate to A, then the pair $(\mathfrak{F}, \mathfrak{F}')$ is called *adjacent pair of families of groups*. In the case that A is normal in G a $(\mathfrak{F}, \mathfrak{F}')$-free geometric unitary bordism class $[M, \partial M, f]$ of X is equivalent to $\sum_{j=1}^{l} [U_j, \partial U_j, f|_{U_j}]$ where the U_j's are disjoint G-equivariant tubular neighborhoods of the M_j^A's and these sets are the connected components of the A-fixed point set M^A. Since the normal bundle of the fixed point set M_j^A may be classified with a map to an appropriate classifying space, the groups $\Omega_*^G\{\mathfrak{F}, \mathfrak{F}'\}(X)$ become isormophic to the direct sum of G/A-free equivariant unitary bordism groups of the prroduct of X^A with an appropriate classifying space (see Ángel, Gómez, and B. Uribe [n.d., Thm. 4.5]). To introduce this result we need to understand how the fixed points of universal equivariant bundles behave. This is the subject of the next section.

2 Equivariant vector bundles and fixed points

2.1 Complex representations. Let G be a compact Lie group and A a closed and normal subgroup of G fitting in the exact sequence

$$1 \to A \to G \to Q \to 1.$$

Let $\rho : A \to U(V_\rho)$ be an irreducible unitary representation of A, denote by $\mathrm{Irr}(A)$ the set of isomorphism classes of irreducible representations of A and let W be a finite dimensional complex G-representation. Then we have an isomorphism of A-representations

$$\bigoplus_{\rho \in \mathrm{Irr}(A)} V_\rho \otimes \mathrm{Hom}_A(V_\rho, W) \xrightarrow{\cong} W.$$

The group G acts on the set of A-representations

$$(g \cdot \rho)(a) := \rho(g^{-1} a g)$$

and therefore it acts on $\mathrm{Irr}(A)$. Denote by $G_\rho := \{g \in G | g \cdot \rho \cong \rho\}$ the isotropy group of the isomorphism class of ρ and denote $Q_\rho := G_\rho / A$.

If $g \cdot \rho \cong \rho$ then there exists $M \in U(V_\rho)$ such that $g \cdot \rho(a) = M^{-1}\rho(a)M$. Since this matrix M is unique up to a central element, we obtain a homomorphism $f : G_\rho \to PU(V_\rho)$ which fits in the following diagram

$$\begin{array}{ccc} A & \overset{\iota}{\hookrightarrow} & G_\rho \\ {\scriptstyle\rho}\downarrow & & \downarrow{\scriptstyle f} \\ U(V_\rho) & \overset{p}{\longrightarrow} & PU(V_\rho). \end{array}$$

thus making V_ρ into a projective G_ρ-representation.

Define the $\mathbb{S}^1$-central extension $\widetilde{G}_\rho := f^*U(V_\rho)$ of G_ρ which fits into following diagram

$$\begin{array}{ccccc} & & \mathbb{S}^1 & & \mathbb{S}^1 \\ & & \downarrow & & \downarrow \\ A & \overset{\widetilde{\iota}}{\longrightarrow} & \widetilde{G}_\rho & \overset{\widetilde{f}}{\longrightarrow} & U(V_\rho) \\ \downarrow{\scriptstyle =} & & \downarrow & & \downarrow \\ A & \overset{\iota}{\longrightarrow} & G_\rho & \overset{f}{\longrightarrow} & PU(V_\rho). \end{array}$$

endowing V_ρ with the structure of a $\widetilde{G}_\rho$-representation where $\mathbb{S}^1$ acts by multiplication with scalars.

The vector space $\mathrm{Hom}_A(V_\rho, W)$ is also a $\widetilde{G}_\rho$-representation where for $\phi \in \mathrm{Hom}_A(V_\rho, W)$ and $\widetilde{g} \in \widetilde{G}_\rho$ we set

$$(\widetilde{g} \bullet \phi)(v) := g\phi(\widetilde{f}(\widetilde{g})^{-1}v).$$

It follows that A acts trivially on $\mathrm{Hom}_A(V_\rho, W)$ and moreover the lements of $\mathbb{S}^1$ act by multiplication of their inverse.

Hence $V_\rho \otimes \mathrm{Hom}_A(V_\rho, W)$ is a G_ρ representation, where $\mathrm{Hom}_A(V_\rho, W)$ is a $\widetilde{Q}_\rho := \widetilde{G}_\rho / A$ representation where $\mathbb{S}^1$ acts by multiplication of the inverse. Here $\widetilde{Q}_\rho$ is an $\mathbb{S}^1$-central extension of Q_ρ.

Since the isotropy group G_ρ contains the connected component of the identity in G, the index $[G\colon G_\rho]$ is finite and we may induce the G_ρ-representation $V_\rho \otimes \mathrm{Hom}_A(V_\rho, W)$ to G thus obtaining the following result.

Theorem 2.1. *There is a canonical isomorphism of G-representations*

$$\bigoplus_{\rho \in G \backslash \mathrm{Irr}(A)} Ind_{G_\rho}^{G} \left(V_\rho \otimes \mathrm{Hom}_A(V_\rho, W)\right) \cong W$$

where ρ runs over representatives of the orbits of the action of G on $Irr(A)$.

2.2 Equivariant complex bundles. The previous result generalizes to equivariant complex vector bundles, but prior to showing this generalization we need to recall the multiplicative induction map introduced in Bix and tom Dieck [1978, §4]. Let H be a closed subgroup of the compact Lie group G. The right adjoint to the restriction functor r_H^G from G-spaces to H-spaces is called the multiplicative induction functor and takes an H-space Y and returns the G-space

$$m_H^G(Y) := \mathrm{map}(G, Y)^H$$

of H-equivariant maps from G to Y, with G considered as an H-space by left multiplication. The G-action on $m_H^G(Y)$ is given by $(g \cdot f)(k) := f(kg)$, it is homeomorphic to the space of sections of the projection map $G \times_H Y \to G/H$ and, in the case that G/H is finite, it is homeomorphic to $[G : H]$ copies of Y.

There is a homeomorphism

$$\mathrm{map}(X, m_H^G(Y))^G \xrightarrow{\cong} \mathrm{map}(r_H^G(X), Y)^H, \quad F \mapsto (x \mapsto F(x)(1_G))$$

whose inverse maps f to $m_H^G(f) \circ p_H^G$ where $p_H^G : X \to m_H^G(r_H^G(X))$, $p_H^G(x)(g) = gx$, is the unit of the adjunction.

Now consider a G-space X on which the closed and normal subgroup A acts trivially. Take a G-equivariant complex vector bundle $p : E \to X$ and assume that E has an hermitian metric in such a way that G acts through unitary matrices on the complex fibers. For a complex A-respresentation $\rho : A \to U(V_\rho)$ denote by $\mathbb{V}_\tau$ the trivial A-vector bundle $\pi_2 : V_\rho \times X \to X$.

The complex vector bundle $\mathrm{Hom}_A(\mathbb{V}_\rho, E)$ is a $\widetilde{Q}_\rho$-equivariant complex vector bundle where $\mathbb{S}^1$-acts on the fibers bymultiplication of the inverse, $\mathbb{V}_\rho \otimes \mathrm{Hom}_A(\mathbb{V}_\rho, E)$ is a G_ρ-equivariant complex vector bundle and

$$(p_{G_\rho}^G)^* \left(m_{G_\rho}^G (\mathbb{V}_\rho \otimes \mathrm{Hom}_A(\mathbb{V}_\rho, E)) \right) \to X$$

is a G-equivariant complex vector bundle over X.

Theorem 2.2. *Ángel, Gómez, and B. Uribe [n.d., Thm. 2.7] Let G be a compact Lie group, A a closed and normal subgroup, X a G-space on which A acts trivially and $E \to X$ a G-equivariant complex vector bundle. Then there is an isomorphism of G-equivariant complex vector bundles*

$$\bigoplus_{\rho \in G \backslash \mathrm{Irr}(A)} (p_{G_\rho}^G)^* \left(m_{G_\rho}^G (\mathbb{V}_\rho \otimes \mathrm{Hom}_A(\mathbb{V}_\rho, E)) \right) \xrightarrow{\cong} E$$

where ρ runs over representatives of the orbits of the G-action on the set of isomorphism classes of A-irreducible representations.

With the same hypothesis as in the previous theorem, there is an induced decomposition in equivariant K-theory

$$K^*_G(X) \cong \bigoplus_{\rho\in G\backslash \mathrm{Irr}(A)} {}^{\widetilde{Q}_\rho}K^*_{Q_\rho}(X), \quad E \mapsto \bigoplus_{\rho\in G\backslash \mathrm{Irr}(A)} \mathrm{Hom}_A(\mathbb{V}_\rho, E)$$

where ${}^{\widetilde{Q}_\rho}K^*_{Q_\rho}(X)$ is the $\widetilde{Q}_\rho$-twisted Q_ρ- equivariant K-theory of X which is built out of the Grothendieck group of $\widetilde{Q}_\rho$-equivariant complex vector bundles over X on which the central $\mathbb{S}^1$ acts by multilpication of the fibers.

2.3 Classifying spaces. The decomposition described above can also be written at the level of classifying spaces; let us set up the notation first. Let G be a compact Lie group and $\widetilde{G}$ a $\mathbb{S}^1$-central group extension of G. Let $\widetilde{\mathbf{C}}^\infty$ be a countable direct sum of all complex irreducible $\widetilde{G}$ representations on which $\mathbb{S}^1$ acts by multiplication of their inverse. Denote by ${}^{\widetilde{G}}B_GU(n)$ the Grasmannian of n-planes of $\widetilde{\mathbf{C}}^\infty$ and denote by ${}^{\widetilde{G}}\gamma_GU(n)$ the canonical n-plane bundle over ${}^{\widetilde{G}}B_GU(n)$. The complex vector bundle

$$\mathbb{C}^n \to {}^{\widetilde{G}}\gamma_GU(n) \to {}^{\widetilde{G}}B_GU(n)$$

is a universal $\widetilde{G}$-twisted G-equivariant complex vector bundle of rank n. Denote by $\gamma_GU(n) \to B_GU(n)$ the universal G-equivariant complex vector bundle of rank n.

Take a closed subgroup A of G, let N_A denote the normalizer of A in G and $W_A := N_A/A$. Consider the fixed point set $B_GU(n)^A$ and the restriction $\gamma_GU(n)|_{B_GU(n)^A}$ of the universal bundle to this fixed point set. Take $\rho \in \mathrm{Irr}(A)$ and by the arguments above we have that

$$\mathrm{Hom}_A(\mathbb{V}_\rho, \gamma_GU(n)|_{B_GU(n)^A})$$

is a $(\widetilde{W_A})_\rho$-twisted $(W_A)_\rho$-equivariant complex bundle, but since the space $B_GU(n)^A$ is not necessarily connected, it may not have constant rank. Therefore Theorem 2.2 implies the following equivariant homotopy equivalence.

Theorem 2.3. *Ángel, Gómez, and B. Uribe [n.d., Thms. 3.3 & 3, 5] There are W_A-equivariant homotopy equivalences*

$$\bigsqcup_{n=0}^{\infty} \gamma_GU(n)^A \simeq \left(\bigsqcup_{n=0}^{\infty} \gamma_{W_A}U(n_1)\right) \times \prod_{\substack{\rho\in W_A\backslash \mathrm{Irr}(A)\\ \rho\neq 1}} m^{W_A}_{(W_A)_\rho}\left(\bigsqcup_{n_\rho=0}^{\infty} {}^{(\widetilde{W_A})_\rho}B_{(W_A)_\rho}U(n_\rho)\right),$$

$$\bigsqcup_{n=0}^{\infty} B_GU(n)^A \simeq \prod_{\rho\in W_A\backslash \mathrm{Irr}(A)} m^{W_A}_{(W_A)_\rho}\left(\bigsqcup_{n_\rho=0}^{\infty} {}^{(\widetilde{W_A})_\rho}B_{(W_A)_\rho}U(n_\rho)\right).$$

If G is abelian then A is normal, G acts trivially on $\mathrm{Irr}(A)$ and all the irreducible representations are 1-dimensional. Therefore we get a G/A-homotopy equivalence

$$\gamma_G U(n)^A \simeq \coprod_{\substack{(n_\rho)_{\rho\in\mathrm{Irr}(A)}\\ \sum_\rho n_\rho = n}} \left(\gamma_{G/A} U(n_1) \times \prod_{\substack{\rho\in\mathrm{Irr}(A)\\ \rho\neq 1}} B_{G/A} U(n_\rho) \right).$$

In order to get a similar formula for the case on which G is not abelian we need to add up some notation and make some choices. Let $\mathcal{P}(n,A)$ be the set of arrangements of non-negative integers $(n_\rho)_{\rho\in\mathrm{Irr}(A)}$ such that

$$\sum_{\rho\in\mathrm{Irr}(A)} n_\rho|\rho| = n,$$

then non-equivariantly there is a homotopy equivalence

$$B_G U(n)^A \simeq \coprod_{(n_\rho)\in\mathcal{P}(n,A)} \prod_{\rho\in\mathrm{Irr}(A)} \left({}^{(\widetilde{W_A})_\rho} B_{(W_A)_\rho} U(n_\rho) \right).$$

The group W_A acts on $\mathcal{P}(n,A)$ on the right permuting the arrangements, i.e. the action of $g \in W_A$ on the arrangement (n_ρ) is the arrangement $(n_\rho)\cdot g := (n_{g\cdot\rho})$ meaning that it has the number $n_{g\cdot\rho}$ in the coordinate ρ. Denote by $(W_A)_{(n_\rho)}$ the isotropy group of the arrangement (n_ρ). Rearranging the terms we obtain the following W_A-equivariant homotopy equivalence

(2.4) $$B_G U(n)^A \simeq$$

$$\coprod_{(n_\rho)\in\mathcal{P}(n,A)/W_A} W_A \underset{(W_A)_{(n_\rho)}}{\times} \left(\prod_{\rho\in(W_A)_{(n_\rho)}\backslash \mathrm{Irr}(A)} m^{(W_A)_{(n_\rho)}}_{(W_A)_\rho\cap(W_A)_{(n_\rho)}} \left({}^{(\widetilde{W_A})_\rho} B_{(W_A)_\rho} U(n_\rho) \right) \right)$$

where (r_ρ) runs over representatives of the orbits of the action of W_A on $\mathcal{P}(n,A)$, and ρ runs over representatives of the orbits of the action of $(W_A)_{(n_\rho)}$ on $\mathrm{Irr}(A)$.

For the calculation of the equivariant unitary bordism of adjacent families of groups we need to consider only the arrangements of non-negative integers (n_ρ) such that the number associated to the trivial representation is zero, i.e. $n_1 = 0$. Denote by $\overline{\mathcal{P}}(n,A)$ the set of arrangements (n_ρ) such that $n_1 = 0$ and define the W_A space:

(2.5) $$C_{N_A,A}(k) :=$$

$$\coprod_{(n_\rho)\in\overline{\mathcal{P}}(k,A)/W_A} W_A \underset{(W_A)_{(n_\rho)}}{\times} \left(\prod_{\substack{\rho\in(W_A)_{(n_\rho)}\backslash \mathrm{Irr}(A)\\ \rho\neq 1}} m^{(W_A)_{(n_\rho)}}_{(W_A)_\rho\cap(W_A)_{(n_\rho)}} \left({}^{(\widetilde{W_A})_\rho} B_{(W_A)_\rho} U(n_\rho) \right) \right)$$

Therefore we have the following W_A-homotopy equivalence

$$\gamma_G U(n)^A \simeq \bigsqcup_{k=0}^{n} \gamma_{W_A} U(n-k) \times C_{N_A,A}(k) \tag{2.6}$$

such that in the case that G is abelian we have the simple formula

$$C_{G,A}(k) = \bigsqcup_{(n_\rho)\in\overline{\mathcal{P}}(k,A)} \prod_{\substack{\rho\in\mathrm{Irr}(A)\\ \rho\neq 1}} B_{G/A} U(n_\rho). \tag{2.7}$$

Now we are ready to state the relation between the G-equivariant unitary bordism groups of adjacent pair of families of groups and the classifying spaces defined above.

Theorem 2.8. *Ángel, Gómez, and B. Uribe [n.d., Cor. 4.6] Let G be a finite group, X a G-space and $(\mathfrak{F}, \mathfrak{F}')$ an adjacent pair of families differing by the conjugacy class of the subgroup A, then there is an isomorphism*

$$\Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X) \cong \bigoplus_{0\leq 2k\leq n} \Omega_{n-2k}^{W_A}\{\{1\}\}(X^A \times C_{N_A,A}(k))$$

where $\{1\}$ is the family of subgroups of W_A which only contains the trivial group.

Take a bordism class $[M, \partial M, f : M \to X]$ in $\Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X)$ and note that $M^A \cap M^{gAg^{-1}} = \varnothing$ whenever g does not belong to N_A. Then choose a N_A-equivariant tubular neighborhood U of M^A such that its G-orbit $G\cdot U$ is a G-equivariant tubular neighborhood of $G \cdot M^A$ and such that

$$G \underset{N_A}{\times} U \overset{\cong}{\to} G \cdot U, \quad [(g, u)] \to gu$$

is a G-equivariant diffeomorphism. The assignment

$$[M, \partial M, f : M \to X] \mapsto [U, \partial U, f|_U : U \to X]$$

induces an isomorphism

$$\Omega_n^G\{\mathfrak{F}, \mathfrak{F}'\}(X) \overset{\cong}{\to} \Omega_n^{N_A}\{\mathfrak{F}|_{N_A}, \mathfrak{F}'|_{N_A}\}(X).$$

Let M^A_{n-2k} denote the component of M^A which is a $(n-2k)$-dimensional W_A-free manifold and such that $M^A = \bigcup_{0\leq 2k\leq n} M^A_{n-2k}$. The tubular neighborhood U is N_A-equivariantly diffeomorphic to $\bigcup_{0\leq 2k\leq n} D(\nu_{n-2k})$ where $\nu_{n-2k} \to M^A_{n-2k}$ is the normal bundle of the inclusion $M^A_{n-2k} \to M$. Since the trivial A-representation does not appear on the fibers of the normal bundles, by Theorem 2.3 and formula (2.6) we know that the bundle ν_{n-2k} is classified by a W_A-equivariant map $h_{n-2k} : M^A_{n-2k} \to C_{N_A,A}(k)$. The bordism class $[M^A_{n-2k}, f|_{M^A_{n-2k}} \times h_{n-2k} : M^A_{n-2k} \to X^A \times C_{N_A,A}(k)]$ belongs to $\Omega^{W_A}_{n-2k}\{\{1\}\}(X^A \times C_{N_A,A}(k))$ and the assignment

$$[M, \partial M, f : M \to X] \mapsto \bigoplus_{0 \le 2k \le n} [M^A_{n-2k}, f|_{M^A_{n-2k}} \times h_{n-2k} : M^A_{n-2k} \to X^A \times C_{N_A,A}(k)]$$

induces the desired isomorphism.

3 The evenness conjecture for finite abelian groups

In this section we will outline the main ingredients that used Landweber [1972] in the cyclic group case, Stong [1970] in the p-group case and Ossa [1972] in the general case to show that evenness conjecture holds for finite abelian groups. The conjecture also holds for compact abelian groups, Löffler [1974] showed it for the homotopic G-equivariant unitary bordism groups in the case that G is a unitary torus, and Comezaña in May [1996, §XXVIII] generalized it to any compact abelian group. Comezaña furthermore showed that the map from the G-equivariant unitary bordism groups to the homotopic ones is injective whenever G is compact abelian thus proving the evenness conjecture for any compact abelian group. In this work we will address the finite group case.

Prior to addressing the study of the G-equivariant unitary bordism groups for finite abelian groups we need to recall some results on the unitary bordism groups.

Thom's remarkable Theorem 1954 shows that the unitary bordism groups Ω_* can be calculated as the stable homotopy groups $\lim_k \pi_{n+k}(MU(k))$ of the Thom spaces $MU(k)$ of the canonical complex vector bundles over $BU(k)$. Milnor [1960, Thm. 3] showed that the these stable homotopy groups are zero if n is odd and free abelian if n is even with a number of generators equal to the number of partitions of $n/2$. Indepently Novikov [1960, Thm. 4] showed that as a ring the unitary bordism groups are isomorphic to the ring of polynomials over the integers with generators x_{2i} of degree $2i$ for $i \ge 1$. The spectrum MU that the Thom spaces $MU(k)$ defines permitted Atiyah [1961] to define the homotopy unitary bordism groups $MU_*(X)$ and the homotopy unitary cobordism groups $MU^*(X)$ of a space X as a generalized homology and cohomology theory respectively. Thom's theorem implies that for X a CW-complex the unitary bordism groups over X are equivalent to the homotopic ones $\Omega_*(X) \cong MU_*(X)$ via the Thom-Pontrjagin map.

The spectral sequence of Atiyah and Hirzebruch [1961] (cf. Kochman [1996, §4.2]) applied to the unitary bordism groups of a CW-complex X produces a spectral sequence which converges to $\Omega_*(X)$ and whose second page is $E^2_{p,q} \cong H_p(X; \Omega_q)$; let us call this spectral sequence the *bordism spectral sequence*. The Thom homomorphism

$$\mu : \Omega_*(X) \to H_*(X;\mathbb{Z}), \ \ [M, f : M \to X] \mapsto f_*[M],$$

which takes a unitary bordism element in X and maps it to he image under f of the fundamental class $[M] \in H_*(M;\mathbb{Z})$, is a natural transformation of homology theories and is also the edge homomorphism $\Omega_*(X) \to E^2_{*,0} \cong H_*(X;\mathbb{Z})$ of the spectral sequence.

Whenever X is a CW-complex whose homology $H_*(X;\mathbb{Z})$ is free abelian then by Conner and Smith [1969, Lemma 3.1] the bordism spectral sequence collapses, the unitary bordism group $\Omega_*(X)$ is a free Ω_*-module and the homomorphism induced by Thom map

$$\widetilde{\mu} : \mathbb{Z} \underset{\Omega_*}{\otimes} \Omega_*(X) \to H_*(X;\mathbb{Z})$$

is an isomorphism.

Applying the bordism spectral sequence to the unitary bordism groups of $BU(n)$ it is shown in Kochman [1996, Prop. 4.3.3] that $\Omega_*(BU(n))$ is a free Ω_*-module with basis

$$\Omega_*(BU(n)) \cong \Omega_*\{\alpha_{k_1}\alpha_{k_2}\dots\alpha_{k_n} : k_1 \leq \dots \leq k_n\}$$

where $\alpha_{k_1}\alpha_{k_2}\dots\alpha_{k_n}$ is the unitary bordism class of the bordism element

$$(\mathbb{C}P^{k_1} \times \dots \times \mathbb{C}P^{k_n}, F : \mathbb{C}P^{k_1} \times \dots \times \mathbb{C}P^{k_n} \to BU(n))$$

where the map F classifies the canonical rank n complex vector bundle over the product of projective spaces.

In Conner and Smith [1969, Prop. 3.6] it is shown that if X is a finite CW-complex such that the Thom homomorphism is surjective then the bordism spectral sequence collapses. Whenever BG is the classifying space of a finite group G Landweber [1971, Thm. 3] showed that the following conditions are equivalent:

- The Thom homomorphism $\mu : \Omega_*(BG) \to H_*(BG;\mathbb{Z})$ is surjective.
- The bordism spectral sequence collapses.
- G has periodic cohomology, i.e. every abelian subgroup of G is cyclic.
- $H^n(BG;\mathbb{Z}) = 0$ for all odd n.
- The projective dimension of $\Omega_*(BG)$ as a Ω_*-module is 1 or 0.

The previous result implies that whenever we consider the cyclic group $G = \mathbb{Z}/k$ of order k, the bordism classes $[L^{2n+1}(k), \iota : L^{2n+1}(k) \to B\mathbb{Z}_k]$ of the Lens spaces $L^{2n+1}(k) := S_k^{2n+1}/(\mathbb{Z}/k)$, where S_k^{2n+1} denotes the sphere of unit vectors in $\mathbb{C}^{n+1}$

with the $\mathbb{Z}/k$-action given by multiplication of the root of unity $e^{\frac{2\pi i}{k}}$, generate $\Omega_*(B\mathbb{Z}_k)$ as a Ω_*-module.

One property of finite abelian groups that will be used is the following. If A is a subgroup of a finite abelian group G and Γ is a product of classifying spaces of the form $B_G U(k)$, then by Theorem 2.3 the fixed point set Γ^A is a product of classifying spaces of the form $B_{G/A}U(l)$. This fact permits to use an induction hypothesis when calculating the equivariant unitary bordism groups of products of spaces of the form $B_G U(k)$.

Now when can start with the proof of the evenness conjecture for finite abelian groups. First we will handle the case of cyclic p-groups following Landweber [1972], then we will review the case of general abelian p-groups following Stong [1970] and we will show the general case using a simple argument on localization shown in Ossa [1972].

3.1 Cyclic p-groups. Let G be a cyclic group of order p^s a power of the prime p. Let $\Gamma := \prod_{i=1}^{l} B_G U(k_i)$ be a product of spaces of the form $B_G U(k)$ and $\mathfrak{F}_t = \{H \subset G : |H| \leq p^t\}$ the family of of subgroups or order bounded by p^t; the family $\mathfrak{F}_s$ is the family of all subgroups of G and therefore $\Omega_*^G(\ ) = \Omega_*^G\{\mathfrak{F}_s\}(\ )$. Let us split $\Omega_*^G(\ ) = \Omega_+^G(\ ) \oplus \Omega_-^G(\ )$ where $\Omega_+^G(\ )$ denotes the even degree bordism groups and $\Omega_-^G(\ )$ the odd degree ones. We will prove by induction on the size of the group that for any $0 \leq t < s$ the following properties hold:

- $\Omega_*^G\{\mathfrak{F}_s, \mathfrak{F}_t\}(\Gamma)$ is a free Ω_*-module in even dimensional generators.
- $\Omega_+^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma)$ is a free Ω_*-module.
- The boundary homomorphism is surjective

$$\Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_t\}(\Gamma) \xrightarrow{\partial} \Omega_-^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma).$$

Let us see that these properties imply that $\Omega_*^G(\Gamma)$ is a free Ω_*-module in even dimensional generators. Since $\Omega_*^G\{\mathfrak{F}_s, \mathfrak{F}_0\}(\Gamma)$ is a free Ω_*-module in even dimensional generators, the long exact sequence associated to the families of groups $\mathfrak{F}_0 \subset \mathfrak{F}_s$ induce the exact sequence

$$0 \to \Omega_+^G\{\mathfrak{F}_0\}(\Gamma) \to \Omega_+^G(\Gamma) \to \Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_0\}(\Gamma) \xrightarrow{\partial} \Omega_-^G\{\mathfrak{F}_0\}(\Gamma) \to \Omega_-^G(\Gamma) \to 0.$$

The unitary bordism group of free actions $\Omega_*^G\{\mathfrak{F}_0\}(\Gamma)$ is isomorphic to $\Omega_*(BG \times \prod_{i=1}^{l} BU(k_i))$ since both $EG \times B_G U(k_i)$ and $EG \times BU(k_i)$ classify G-equivariant complex vector bundles of rank k_i over free G-spaces. The unitary bordism groups of

$BU(k_i)$ are free Ω_*-modules in even dimensional generators, and therefore by the Kunneth theorem we have that

$$\Omega_*^G\{\mathfrak{F}_0\}(\Gamma) \cong \Omega_*(BG) \underset{\Omega_*}{\otimes} \Omega_* \left(\prod_{i=1}^{l} BU(k_i) \right).$$

Hence we have that $\Omega_+^G\{\mathfrak{F}_0\}(\Gamma)$ is a free Ω_*-module in even degrees and that $\Omega_-^G\{\mathfrak{F}_0\}(\Gamma)$ is all p-torsion. Consider a unitary bordism class defined by the map $h : M \to \prod_{i=1}^{l} BU(k_i)$ and denote by $E := E_1 \oplus \cdots \oplus E_l$ with E_j the complex vector bundle that the map $\pi_j \circ h : M \to BU(k_j)$ defines. Take the ball $B_{p^s}^{2n+2}$ of vectors in $\mathbb{C}^{n+1}$ with norm less than 1 endowed with the action of G given by multiplication by $e^{\frac{2\pi i}{p^s}}$ and consider the G-equivariant $\prod_{i=1}^{l} U(k_i)$ complex bundle that the product $B_{p^s}^{2n+2} \times E \to B_{p^s}^{2n+2} \times M$ defines.

This G-equivariant $\prod_{i=1}^{l} U(k_i)$ complex bundle is classified by a G-equivariant map

$$f : B_{p^s}^{2n+2} \times M \to \Gamma$$

and its G-equivarian unitary bordism class $[B_{p^s}^{2n+2} \times M, f]$ belongs to $\Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_0\}(\Gamma)$. Its boundary is $[S_{p^s}^{2n+1} \times M, f|_{S_{p^s}^{2n+1} \times M}]$ and it belongs to $\Omega_-^G\{\mathfrak{F}_0\}(\Gamma)$. By the Kunneth isomorphism described above we know that the unitary bordism classes $[S_{p^s}^{2n+1} \times M, f|_{S_{p^s}^{2n+1} \times M}]$ generate $\Omega_-^G\{\mathfrak{F}_0\}(\Gamma)$ and therefore the boundary homomorphism $\Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_0\}(\Gamma) \xrightarrow{\partial} \Omega_-^G\{\mathfrak{F}_0\}(\Gamma)$ is surjective. This implies that $\Omega_-^G(\Gamma)$ is trivial.

Since $\Omega_+^G\{\mathfrak{F}_0\}(\Gamma) \cong \Omega_+(\prod_{i=1}^{l} BU(k_i))$ is a free Ω_*-module, and by hypothesis $\Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_0\}$ also, then it implies that $\Omega_+^G(\Gamma)$ is a free Ω_*-module.

In particular we have that the G-equivariant unitary bordism group Ω_*^G is a free Ω_*-module in even dimensional generators.

Now let us sketch the proof of the properties cited above. Let us assume that the properties hold for cyclic groups of order less than p^s and let us proceed by induction on the families of subgroups of G. For the adjacent pair of families $(\mathfrak{F}_s, \mathfrak{F}_{s-1})$ differing by the group G, we know by Theorem 2.8 that $\Omega_*^G\{\mathfrak{F}_s, \mathfrak{F}_{s-1}\}(\Gamma)$ is a direct sum of groups $\Omega_*(\Gamma^G \times \Gamma')$ where both Γ^G and Γ' are products of classifying spaces of unitary groups. Therefore $\Omega_*^G\{\mathfrak{F}_s, \mathfrak{F}_{s-1}\}(\Gamma)$ is a free Ω_*-module in even dimensional generators and we have started our induction.

Now let us assume that the properties hod for the pair of families $(\mathfrak{F}_s, \mathfrak{F}_j)$ for $s > j \geq t$. Therefore we get the following exact sequence of groups

$$\begin{aligned} 0 \to \Omega_+^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma) \to \Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_{t-1}\}(\Gamma) \to \Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_t\}(\Gamma) \xrightarrow{\partial} \\ \Omega_-^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma) \to \Omega_-^G\{\mathfrak{F}_s, \mathfrak{F}_{t-1}\}(\Gamma) \to 0. \end{aligned} \tag{3.1}$$

Since the pair of families $(\mathfrak{F}_t, \mathfrak{F}_{t-1})$ differ by the cyclic group H or order p^t, G/H is a cyclic group of order p^{s-t}, and Γ^H is a product of classifying spaces of the form $B_{G/H}U(k)$, then by Theorem 2.8 there is an isomorphism

$$\Omega_*^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma) \cong \bigoplus_{k\geq 0} \Omega_{*-2k}^{G/H}\{\mathfrak{F}_0\}(\Gamma^H \times C_{G,H}(k))$$

where both Γ^H and $C_{G,H}(k)$ are disjoint unions of products of classifying spaces of the form $B_{G/H}U(k)$.

Therefore we know that $\Omega_+^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma)$ is a free Ω_*-module and by the induction hypothesis we know that the boundary map

$$\Omega_+^{G/H}\{\mathfrak{F}_{s-t}, \mathfrak{F}_0\}(\Gamma^H \times C_{G,H}(k)) \xrightarrow{\partial} \Omega_-^{G/H}\{\mathfrak{F}_0\}(\Gamma^H \times C_{G,H}(k)) \tag{3.2}$$

is surjective. A bordism class in $\Omega_-^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma)$ can be represented by a class $[D(E), f : D(E) \to \Gamma]$ where $D(E)$ is the disk bundle of a G-equivariant vector bundle $E \to M$ over a manifold M on which H acts trivially and G/H acts freely, and such that the trivial representation of H does not appear on the fibers of E. This bundle is classified by a G/H-equivariant map $h : M \to C_{G,H}(k)$ for some k, and the bordism class $[M, f|_M \times h : M \to \Gamma^H \times C_{G,H}(k)]$ lives in $\Omega_-^{G/H}\{\mathfrak{F}_0\}(\Gamma^H \times C_{G,H}(k))$. By the surjectivicty of (3.2) there is a bordism class $[Z, F \times \tilde{h} : Z \to \Gamma^H \times C_{G,H}(k)]$ in $\Omega_+^{G/H}\{\mathfrak{F}_{s-t}, \mathfrak{F}_0\}(\Gamma^H \times C_{G,H}(k))$ such that $\partial Z = M$, $F|_M = f|_M$ and $\tilde{h}|_M = h$. Let $p : V \to Z$ denote the G-equivariant vector bundle over Z that the map $\tilde{h}$ defines and note that the bordism class $[D(V), F \circ p : D(V) \to \Gamma]$ defines an element in $\Omega_*^G\{\mathfrak{F}_s, \mathfrak{F}_t\}(\Gamma)$ since the trivial H-representation does not appear on the fibers of V and the action of G/H over M is free. The boundary of $D(V)$ is the union of the sphere bundle $S(V)$ and $D(V)|_M = D(E)$, but since $S(V)$ is $\mathfrak{F}_{t-1}$-free we have that

$$\begin{aligned}\partial[D(V), F \circ p : D(V) \to \Gamma] &= [D(E), p|_E \circ f|_M : D(E) \to \Gamma] \\ &= [D(E), f : D(E) \to \Gamma]\end{aligned}$$

and therefore the boundary map

$$\Omega_+^G\{\mathfrak{F}_s, \mathfrak{F}_t\}(\Gamma) \xrightarrow{\partial} \Omega_-^G\{\mathfrak{F}_t, \mathfrak{F}_{t-1}\}(\Gamma)$$

is surjective.

By the long exact sequence of (3.1) we deduce that $\Omega_*^G\{\mathfrak{F}_s, \mathfrak{F}_{t-1}\}(\Gamma)$ is a free Ω_*-module on even-dimensional generators and we conclude that the properties also hold for the pair of families $(\mathfrak{F}_s, \mathfrak{F}_{t-1})$.

Therefore the evenness conjecture holds for cyclic p-groups Landweber [1972, Thm. 1’].

3.2 General p-groups. The argument to show the evenness conjecture for general p-groups is more elaborate than the one done above for cyclic p-groups. We will follow the original proof of Stong [1970] on which the author uses very cleverly the Thom isomorphism and the long exact sequence for pairs of spaces in order to understand the long exact sequence for a pair of families once restricted to a special kind of actions on manifolds. Here we shorten the original proof and we highlight its main ingredients.

Let $G = H \times \mathbb{Z}/q$ with $q = p^s$ such that all elements in H have order less or equal than p^s and let

$$\Gamma := \prod_{i=1}^{l} B_G U(k_i)$$

be a product of spaces of the form $B_G U(k)$. We will show by induction on the order of the group G that the bordism group $\Omega_*^G(\Gamma)$ is a free Ω_*-module on even dimensional generators. Therefore let us assume that $\Omega_*^K(\Gamma')$ is a free Ω_*-module in even dimensional generators for all p-groups of order less than the order of G and Γ' any product of classifying spaces of the form $B_K U(l)$.

Following the notation of Stong [ibid.] let us consider the following families of subgroups of G:

- $\mathfrak{F}_a$ is the family of all subgroups of G,
- $\mathfrak{F}_s$ is the family of subgroups whose intersection with $\mathbb{Z}/q$ is proper, i.e. $\mathfrak{F}_s := \{W \subset H \times Z/q \colon \{1\} \times Z/q \not\subset W\}$,
- $\mathfrak{F}_f$ is the family of subgroups whose intersection with $\mathbb{Z}/q$ is the unit subgroup, i.e. $\mathfrak{F}_f := \{W \subset H \times Z/q \colon \{1\} \times Z/q \cap W = \{(1,1)\}$

A manifold M is $\mathfrak{F}_s$-free if for every $x \in M$ the isotropy group $(\mathbb{Z}/q)_x \neq \mathbb{Z}/q$, and it is $\mathfrak{F}_f$-free if the restriction of the action to $\mathbb{Z}/q$ is free.

The classifying space $E\mathfrak{F}_f$ has a free $\mathbb{Z}/q$-action and can be understood as the universal H-equivariant $\mathbb{Z}/q$-principal bundle $E_H\mathbb{Z}/q$ Lück and Uribe [2014, Thm 11.4] . Hence $E\mathfrak{F}_f = E_H\mathbb{Z}/q$ and its quotient $E\mathfrak{F}_f/(\mathbb{Z}/q) = B_H\mathbb{Z}/q$ is the classifynig space of H-equivariant $\mathbb{Z}/q$-principal bundles. By the isomorphism of (1.2), and since the action of $\mathbb{Z}/q$ is free, we get the following isomorphisms (see Stong [1970, Prop. 3.1]):

$$\Omega_*^G\{\mathfrak{F}_f\}(X) \cong \Omega_*^G(X \times E_H\mathbb{Z}/q) \cong \Omega_*^H(X \times_{\mathbb{Z}/q} E_H\mathbb{Z}/q). \tag{3.3}$$

Since both spaces $E_H\mathbb{Z}/q \times B_G U(k_i)$ and $E_H\mathbb{Z}/q \times B_H U(k_i)$ classify $H \times \mathbb{Z}/q$-equivariant $U(k_i)$-principal bundles over spaces with free $\mathbb{Z}/q$-action, we may take the maps

$$B_H U(k_i) \to B_G U(k_i) \to B_H U(k_i),$$

where the left hand side map classifies the G-equivariant complex bundles such that the action of $\mathbb{Z}/q$ is trivial over the total space of the bundle, and the right hand side is the one that forgets the $\mathbb{Z}/q$-action, thus producing G-equivariant homotopy equivalences.

$$E_H\mathbb{Z}/q \times E_H U(k_i) \xrightarrow{\simeq} E_H\mathbb{Z}/q \times E_G U(k_i) \xrightarrow{\simeq} E_H\mathbb{Z}/q \times E_H U(k_i).$$

If we denote by

$$\Gamma' := \prod_{i=1}^{l} B_H U(k_i)$$

and the map $\iota : \Gamma' \to \Gamma$ is the one that classifies trivial $\mathbb{Z}/q$-bundles over H-spaces, then the argument above implies that the following isomorphism holds (see Stong [ibid., Prop. 3.2]) :

(3.4) $$\Omega_*^H(B_H\mathbb{Z}/q \times \Gamma') \cong \Omega_*^G(E_H\mathbb{Z}/q \times \Gamma') \underset{\cong}{\xrightarrow{\iota_*}} \Omega_*^G\{\mathfrak{F}_f\}(\Gamma).$$

Let T be the generator of the group $\mathbb{Z}/q$ and denote by $\mathbb{Z}/p^t$ the subgroup generated by $T^{p^{s-t}}$. A manifold is $\mathfrak{F}_f$-free if and only if $T^{p^{s-1}}$ acts freely and therefore a $(\mathfrak{F}_a, \mathfrak{F}_f)$-manifold M can be localized to the normal bundle of the fixed point set $M^{\mathbb{Z}/p}$ of the subgroup $\mathbb{Z}/p$. The normal bundle is a G-equivariant complex bundle over the trivial $\mathbb{Z}/p$ space and once it is classified to the appropriate spaces $C_{G,\mathbb{Z}/p}(k)$ of (2.5) we obtain the isomorphism (see Stong [ibid., Prop. 3.4]):

(3.5) $$\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(X) \cong \bigoplus_{0\le 2k\le *} \Omega_{*-2k}^{G/(\mathbb{Z}/p)}\left(X^{\mathbb{Z}/p} \times C_{G,\mathbb{Z}/p}(k)\right).$$

Applying the previous isomorphism to $\Gamma = \prod_{i=1}^{l} B_G U(k_i)$ we obtain that

$$\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma)$$

is a free Ω_*-module in even dimensional generators since both $\Gamma^{\mathbb{Z}/p}$ and $C_{G,\mathbb{Z}/p}(k)$ are products of spaces of the form $B_{G/(\mathbb{Z}/p)}U(l)$ and by induction we assumed that the evenness conjecture was true for groups of order less than the one of G and spaces of this type. Therefore the long exact sequence for the pair of families $(\mathfrak{F}_a, \mathfrak{F}_f)$ becomes:

(3.6) $$0 \to \Omega_+^G\{\mathfrak{F}_f\}(\Gamma) \to \Omega_+^G(\Gamma) \to \Omega_+^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma) \xrightarrow{\partial} \Omega_-^G\{\mathfrak{F}_f\}(\Gamma) \to \Omega_-^G(\Gamma) \to 0.$$

A $(\mathfrak{F}_a, \mathfrak{F}_s)$-free manifold M once restricted to the action of $\mathbb{Z}/q$ becomes a $\mathbb{Z}/q$-manifold on which the boundary has no fixed points of the whole group. Therefore the manifold can be localized on the normal bundle of the fixed point set $M^{\mathbb{Z}/q}$ and the information of th normal bundle can be recorded with appropriate maps to the classifying spaces $C_{G,\mathbb{Z}/q}(k)$ of (2.5). Following the same proof as in Theorem 2.8 one obtains the following isomorphism (see Stong [1970, Prop. 3.3]):

$$\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(X) \cong \bigoplus_{0\leq 2k\leq *} \Omega^H_{*-2k}\left(X^{\mathbb{Z}/q} \times C_{G,\mathbb{Z}/q}(k)\right). \tag{3.7}$$

Since both $\Gamma^{\mathbb{Z}/q}$ and $C_{G,\mathbb{Z}/q}(k)$ are products of spaces of the form $B_H U(l)$, by the induction hypothesis we obtain that $\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(\Gamma)$ is a free Ω_*-module in even dimensional generators.

In order to understand the image of the boundary map of (3.6) Stong restricted the equivariant bordism groups to manifolds with a special type of G action. Stong noticed that the image of the boundary map could be determined by restricting to manifolds on which the $\mathbb{Z}/q$-fixed points are of codimension 2 and therefore he studied the class of *special G actions*.

Definition 3.8. Let $G = H \times \mathbb{Z}/q$ be a finite abelian group. The class of *special G actions* is the collection of G-equivariant unitary manifolds M satisfying:

- The restriction to a $\mathbb{Z}/q$-action is semi-free, i.e. for each $x \in M$ the isiotropy group $(\mathbb{Z}/q)_x$ is either $\mathbb{Z}/q$ or $\{1\}$.
- The set $M^{\mathbb{Z}/q}$ of fixed point sets has codimension 2 in M and $\mathbb{Z}/q$ acts in the normal bundle of $M^{\mathbb{Z}/q}$ so that the generator T of $\mathbb{Z}/q$ acts by multiplication by $e^{\frac{2\pi i}{q}}$, or the fixed point set $M^{\mathbb{Z}/q}$ is empty.

The class of special G actions is sufficiently large to permit all constructions done in Section 1, and for a pair of families $(\mathfrak{F}, \mathfrak{F}')$ in G we denote by $\overline{\Omega}_*^G\{\mathfrak{F}, \mathfrak{F}'\}$ the equivariant homology theory defined by using only special G actions. The inclusion of special G actions in the class of all G actions defines natural transformations of homology theories

$$I_* \colon \overline{\Omega}_*^G\{\mathfrak{F}, \mathfrak{F}'\} \to \Omega_*^G\{\mathfrak{F}, \mathfrak{F}'\}$$

preserving the relations between these functors. The G-equivariant unitary bordism groups of special G actions satisfy the following properties:

(i) The natural transformation

$$I_* \colon \overline{\Omega}_*^G\{\mathfrak{F}_f\} \xrightarrow{\cong} \Omega_*^G\{\mathfrak{F}_f\}$$

is an equivalence since every $\mathfrak{F}_f$ action is a special G action.

(ii) The inclusion $(\mathfrak{F}_a, \mathfrak{F}_f) \subset (\mathfrak{F}_a, \mathfrak{F}_s)$ induces an isomorphism

$$\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(X) \xrightarrow{\cong} \overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(X)$$

since $\mathfrak{F}_s$-free special G actions are $\mathfrak{F}_f$-free.

(iii) From the equation (3.7) we get the isomorphism

$$\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(X) \cong \Omega_{*-2}^H(X^{\mathbb{Z}/q} \times B_H U(1)),$$

thus implying that $\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(X)$ maps isomorphically to a direct summand in $\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(X)$.

(iv) For $\Gamma := \prod_{i=1}^l B_G U(k_i)$ the induction hypothesis implies that $\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma)$ is a free Ω_*-module in even dimensional generators. Therefore the canonical maps

$$\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma) \to \Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma) \to \Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_s\}(\Gamma)$$

imply that $\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma)$ also maps isomorphically to a direct summand in $\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma)$.

Let us now concentrate in understanding the five term exact sequence restricted to special G actions

(3.9) $$0 \to \overline{\Omega}_+^G\{\mathfrak{F}_f\}(\Gamma) \to \overline{\Omega}_+^G(\Gamma) \to \overline{\Omega}_+^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma) \xrightarrow{\partial} \overline{\Omega}_-^G\{\mathfrak{F}_f\}(\Gamma) \to \overline{\Omega}_-^G(\Gamma) \to 0.$$

Note that the map $\iota_* : \Gamma' \to \Gamma$ induces the commutative diagram

$$\begin{array}{ccccc} \Omega_{*-2}^H(\Gamma^{\mathbb{Z}/q} \times B_H U(1)) & \xrightarrow{\cong} & \overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma) & \xrightarrow{\partial} & \overline{\Omega}_{*-1}^G\{\mathfrak{F}_f\}(\Gamma) \\ \uparrow & & \uparrow \iota_* & & \uparrow \iota_* \cong \\ \Omega_{*-2}^H(\Gamma' \times B_H U(1)) & \xrightarrow{\cong} & \overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma') & \xrightarrow{\partial} & \overline{\Omega}_{*-1}^G\{\mathfrak{F}_f\}(\Gamma') \end{array}$$

where the middle homomorphism ι_* maps isomorphically $\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma')$ into a direct summand in $\overline{\Omega}_*^G\{\mathfrak{F}_a, \mathfrak{F}_f\}(\Gamma)$ since Γ' is mapped to one connected component of the fixed point set $\Gamma^{\mathbb{Z}/q}$. Therefore the image of the boundary homomorphism ∂ is the same same in both cases.

In what follows we will study the induced boundary homomorphism

(3.10) $$\Omega_{*-2}^H(\Gamma' \times B_H U(1)) \to \overline{\Omega}_{*-1}^G\{\mathfrak{F}_f\}(\Gamma') \cong \Omega_{*-1}^H(\Gamma' \times B_H \mathbb{Z}/q)$$

using the Thom isomorphism, the long exact sequence for pairs and a particular model for $E_H\mathbb{Z}/q$.

Let $\mathbf{C}_H^\infty$ be a countable direct sum of all complex irreducible H-representations and consider the $\mathbb{Z}/q$ action on $\mathbf{C}_H^\infty$ such that the generator T of $\mathbb{Z}/q$ acts by mutliplication of $e^{\frac{2\pi i}{q}}$. The sphere $S(\mathbf{C}_H^\infty)$ of vectors of norm 1 is an $G = \mathbb{Z}/q \times H$ space on which $\mathbb{Z}/q$ acts freely and moreover is a $\mathfrak{F}_f$-free space. Since the non empty fixed point sets are infinitely dimensional spheres we know that this sphere $S(\mathbf{C}_H^\infty)$ is a model for $E_H\mathbb{Z}/q$. The Grassmannian $Gr_1\mathbf{C}_H^\infty$ is a model for $B_HU(1)$ since $\mathbb{Z}/q$ acts trivially on the one dimensional vector spaces, and $\mathbb{Z}/q$ acts on the fibers of the canonical line bundle $\gamma_H U(1) \to B_H U(1)$ by multiplication of $e^{\frac{2\pi i}{q}}$. To simplify the notation denote by

$$\gamma_1 := \gamma_H U(1)$$

and note that $S(\mathbf{C}_H^\infty) \cong S(\gamma_1)$ where $S(\gamma_1)$ denotes the sphere bundle of γ_1.

Consider now the line bundle $\gamma_1^{\otimes q}$ over $B_HU(1)$ which is q-fold tensor product of γ_1. The diagonal map

$$\Delta : \gamma_1 \to \gamma_1^{\otimes q}, \quad v \mapsto v \otimes \cdots \otimes v$$

is a q to 1 map on the fibers of the line bundles and therefore it induces an H-equivariant homeomorphism

$$S(\gamma_1)/(\mathbb{Z}/q) \cong S(\gamma_1^{\otimes q}).$$

Therefore we have that we may take either $S(\gamma_1)/(\mathbb{Z}/q)$ or $S(\gamma_1^{\otimes q})$ as a model for $B_H\mathbb{Z}/q$. The Thom isomorphism

$$\Omega_*^H((D(\gamma_1^{\otimes q}), S(\gamma_1^{\otimes q})) \times \Gamma') \cong \Omega_{*-2}^H(B_HU(1) \times \Gamma')$$

together with the long exact sequence for the pair $(D(\gamma_1^{\otimes q}), S(\gamma_1^{\otimes q}))$ and the induction hypothesis provides a four term exact sequence

$$0 \to \Omega_+^H(\Gamma' \times S(\gamma_1^{\otimes q})) \to \Omega_+^H(\Gamma' \times B_HU(1)) \to \\ \Omega_+^H(\Gamma' \times B_HU(1)) \to \Omega_-^H(\Gamma' \times S(\gamma_1^{\otimes q})) \to 0,$$

where the right hand side homomorphism is precisely the one of (3.10). Therefore we obtain that the boundary homomorphism of (3.10) is surjective, and since by the induction hypothesis $\Omega_+^H(\Gamma' \times B_HU(1))$ is a free Ω_*-module, we conclude that $\Omega_+^H(\Gamma' \times S(\gamma_1^{\otimes q}))$ is also a free Ω_*-module.

Therefore we obtain the following commutative diagram with exact rows

$$\begin{array}{ccccccccccc}
0 \longrightarrow & \Omega_+^G\{\mathfrak{F}_f\}(\Gamma) & \longrightarrow & \Omega_+^G(\Gamma) & \longrightarrow & \Omega_+^G\{\mathfrak{F}_a,\mathfrak{F}_f\}(\Gamma) & \xrightarrow{\partial} & \Omega_-^G\{\mathfrak{F}_f\}(\Gamma) & \longrightarrow 0 \\
& \uparrow\cong & & \uparrow & & \uparrow & & \uparrow\cong & \\
0 \longrightarrow & \overline{\Omega}_+^G\{\mathfrak{F}_f\}(\Gamma) & \longrightarrow & \overline{\Omega}_+^G(\Gamma) & \longrightarrow & \overline{\Omega}_+^G\{\mathfrak{F}_a,\mathfrak{F}_f\}(\Gamma) & \xrightarrow{\partial} & \overline{\Omega}_-^G\{\mathfrak{F}_f\}(\Gamma) & \longrightarrow 0 \\
& \iota_*\uparrow\cong & & \uparrow & & \uparrow & & \iota_*\uparrow\cong & \\
0 \longrightarrow & \overline{\Omega}_+^G\{\mathfrak{F}_f\}(\Gamma') & \longrightarrow & \overline{\Omega}_+^G(\Gamma') & \longrightarrow & \overline{\Omega}_+^G\{\mathfrak{F}_a,\mathfrak{F}_f\}(\Gamma') & \xrightarrow{\partial} & \overline{\Omega}_-^G\{\mathfrak{F}_f\}(\Gamma') & \longrightarrow 0,
\end{array}$$

thus implying that $\Omega_-^G(\Gamma) = 0$ and that $\Omega_+^G(\Gamma)$ is a free Ω_*-module since both $\Omega_+^G\{\mathfrak{F}_f\}(\Gamma)$ and $\Omega_+^G\{\mathfrak{F}_a,\mathfrak{F}_f\}(\Gamma)$ are free Ω_*-modules.

Therefore the evenness conjecture holds for finite abelian p-groups.

3.3 The general case. The proof of the evenness conjecture for general finite abelian groups was done by Ossa [1972] and is based on the proof of Stong for p-groups and appropriate localizations at different primes. For a finite abelian group K denote by $Z_K := \mathbb{Z}[1/|K|]$ the localization of the integers at the ideal generated by the order of K.

Let $G = K \times L$ with K and L finite abelian with $|K|$ and $|L|$ relatively prime and consider the homomorphism $\Omega_*^{K\times L}\{\mathfrak{F}\} \to \Omega_*^L\{\mathfrak{F}\}$ which forgets the K action and $\mathfrak{F}$ is any family of subgroups of L . Let us show that the localized homomorphism

$$\Omega_*^{K\times L}\{\mathfrak{F}\}(\Gamma) \otimes Z_K \to \Omega_*^L\{\mathfrak{F}\}(\Gamma) \otimes Z_K$$

is an isomorphism whenever $\Gamma := \prod_{i=1}^l B_G U(k_i)$. Let us proceed by induction over L and over the family $\{\mathfrak{F}\}$.

For the trivial family $\mathfrak{F} = \{\{1\}\}$ we obtain the isomorphism

$$\Omega_*(BK \times BL \times \prod_i BU(k_i)) \otimes Z_K \xrightarrow{\cong} \Omega_*(BL \times \prod_i BU(k_i)) \otimes Z_K$$

since $\Omega_*(BK) \otimes Z_K \cong \Omega_* \otimes Z_K$.

Whenever the adjacent pair of families $(\mathfrak{F},\mathfrak{F}')$ differ by $H \subset L$ we obtain the homomorphism of long exact sequences

$$\begin{array}{ccccccc}
\longrightarrow & \Omega_*^{K\times L}\{\mathfrak{F}'\}(\Gamma) & \longrightarrow & \Omega_*^{K\times L}\{\mathfrak{F}\}(\Gamma) & \longrightarrow & \Omega_*^{K\times L/H}\{\{1\}\}(\Gamma^H \times \Gamma') & \longrightarrow \\
& \downarrow & & \downarrow & & \downarrow & \\
\longrightarrow & \Omega_*^{L}\{\mathfrak{F}'\}(\Gamma) & \longrightarrow & \Omega_*^{L}\{\mathfrak{F}\}(\Gamma) & \longrightarrow & \Omega_*^{L/H}\{\{1\}\}(\Gamma^H \times \Gamma') & \longrightarrow
\end{array}$$

with Γ' a disjoint union of products of spaces of the form $B_{K\times L/H}U(l)$. Tensoring with Z_K induces an isomorphism on the left vertical arrow by the induction hypothesis on the families and an isomorphism on the right vertical arrow by the induction hypothesis on the group L/H. The 5-lemma implies the desired isomorphism.

Now let $\mathfrak{F}$ be any family of subgroups of K and denote by $\mathfrak{F}\times\Phi$ the family of subgroups of G whose elements are groups $J \times H$ with $J \in \mathfrak{F}$ and H any subgroup of L. Let us show by induction on $\mathfrak{F}$ and on the group K that the localized module

$$\Omega_*^{K\times L}\{\mathfrak{F}\times\Phi\}(\Gamma)\otimes Z_K$$

is a free $\Omega_*\otimes Z_K$-module. Whenever $\mathfrak{F}$ is the trivial family we have shown above that

$$\Omega_*^{K\times L}\{\{1\}\times\Phi\}(\Gamma)\otimes Z_K \xrightarrow{\cong} \Omega^L(\Gamma)\otimes Z_K$$

is an isomorphism.

If the adjacent pair of families $(\mathfrak{F},\mathfrak{F}')$ differ by the subgroup J, then we obtain the long exact sequence

$$\cdots\to\Omega_*^{K\times L}\{\mathfrak{F}'\times\Phi\}(\Gamma)\to\Omega_*^{K\times L}\{\mathfrak{F}\times\Phi\}(\Gamma)\to\Omega_*^{K/J}\{\{1\}\}(\Gamma^{J\times L}\times\Gamma'')\to\cdots$$

where Γ'' is a disjoint union of spaces of the form $B_{K/J}U(l)$. Tensoring with Z_K gives us free $\Omega_*\otimes Z_K$-modules on the left hand side by the induction on families and on the right hand side by the induction on the group K. Therefore the middle term is also a free $\Omega_*\otimes Z_K$-module.

Therefore we have proved that if $\Omega^L(\Gamma)$ is free Ω_*-module then $\Omega^{K\times L}(\Gamma)\otimes Z_K$ is a free $\Omega_*\otimes Z_K$-module. Let us now write $G = P_1\times\cdots\times P_k$ with P_i its sylow p_i-subgroup. Since the evenness conjecture holds for p-groups, we have that $\Omega_*^{P_i}(\Gamma)$ is a free Ω_*-module and therefore $\Omega_*^G(\Gamma)\otimes\mathbb{Z}[1/[G:P_i]]$ is a free $\Omega_*\otimes\mathbb{Z}[1/[G:P_i]]$-module. Since the numbers $[G:P_i]$ are relatively prime it follows that $\Omega_*^G(\Gamma)$ is a free Ω_*-module.

Therefore the evenness conjecture holds for finite abelian groups.

4 The equivariant unitary bordism groups for non abelian groups

The evenness conjecture has been shown to be true for the dihedral groups D_{2p} with p-prime by Ángel, Gómez, and B. Uribe [n.d.], for groups of order pq where p and q are different primes by Lazarov [1972] and for the more general case of metacyclic groups by Rowlett [1980]. In these cases the group G is a semidirect product $\mathbb{Z}/r\rtimes\mathbb{Z}/s$ of cyclic groups with r and s relatively prime, and the study of the equivariant unitary bordism groups is also carried out calculating the equivariant unitary bordism groups of adjacent pair of families of subgroups as is done in the cyclic group case of Section 3.1.

The main tool used by Rowlett to study the metacyclic case is the equivariant unitary spectral sequence constructed by himself in Rowlett [1978, Prop. 2.1]. Suppose that A is a normal subgroup of G and that $Q = G/A$. A family $\mathfrak{F}$ of subgroups of A is called G-invariant if it is closed under conjugation by elements of G. Consider a pair $(\mathfrak{F}, \mathfrak{F}')$ of G-invariant families of A and note that $\Omega_*^A\{\mathfrak{F}, \mathfrak{F}'\}$ becomes a Q-module in the following way. Consider an A-manifold M with action $\theta : A \times M \to M$ and take an element $g \in G$. Define a new action on M by the map $g_*\theta : A \times M \to M$, $g_*(a, m) := \theta(g^{-1}ag, m)$ and denote the action of g on the bordism class $[M, \theta]$ by $\overline{g}[M, \theta] := [M, g_*\theta]$. This action is trivial on elements of A and therefore it boils down to an action of Q. Then there is a first quadrant spectral sequence E^r converging to $\Omega_*^G\{\mathfrak{F}, \mathfrak{F}'\}$ whose second page is

$$E^2_{p,q} \cong H_p(Q, \Omega_q^A\{\mathfrak{F}, \mathfrak{F}'\}).$$

In the case that both groups A and Q are cyclic of relative prime order, the action of Q on $\Omega_+^A\{\mathfrak{F}, \mathfrak{F}'\}$ factors through a permutation action on the free generators and therefore the second page is not difficult to calculate. If we take the family $\mathfrak{F}_A$ of all subgroups of A, the second page of the spectral sequence becomes $H_q(Q, \Omega_q^A)$, and since Ω_*^A is a free Ω-module in even dimensional generators, then we obtain that $\Omega_+^G\{\mathfrak{F}\}$ is a free Ω_*-module. Moreover, the same explicit construction carried out in Section 3.1 can be adopted in this case to show that the long exact sequence associated to the pair of families $\{\mathfrak{F}_a, \mathfrak{F}_A\}$, with $\mathfrak{F}_a$ the family of all subgroups, becomes

$$0 \to \Omega_+^G\{\mathfrak{F}_A\} \to \Omega_+^G \to \Omega_+^G\{\mathfrak{F}_a, \mathfrak{F}_A\} \xrightarrow{\partial} \Omega_-^G\{\mathfrak{F}_A\} \to 0.$$

The same argument as in (3.5) shows that $\Omega_*^G\{\mathfrak{F}_a, \mathfrak{F}_A\}$ is a free Ω_*-module in even dimensional generators and therefore we conclude that Ω_-^G is zero and Ω_+^G a free Ω_*-module.

The spectral sequence defined above can also be used in order to understand the torsion free part of the G-equivariant unitary groups for any abelian group. Take any subgroup A of G and let $(\mathfrak{F}_A, \mathfrak{F}'_A)$ be the adjacent pair of families of G which differ by the conjugacy class of A. Tensoring with the rationals we obtain an isomorphism

$$\Omega_*^G\{\mathfrak{F}_A, \mathfrak{F}'_A\} \otimes \mathbb{Q} \cong \Omega_*^A\{\mathfrak{F}_A, \mathfrak{F}'_A\}^{W_A} \otimes \mathbb{Q}$$

where the right hand side consists of the W_A-invariant part. Since $\Omega_*^A\{\mathfrak{F}_A, \mathfrak{F}'_A\}$ is a free Ω_*-module in even dimensional generators we obtain the isomorphism

$$\Omega_*^G \otimes \mathbb{Q} \cong \bigoplus_{(A)} \Omega_*^A\{\mathfrak{F}_A, \mathfrak{F}'_A\}^{W_A} \otimes \mathbb{Q}$$

where (A) runs over the conjugacy classes of subgroups of G (see Rowlett [ibid., Thm. 1.1], c.f. tom Dieck [1973, Thm. 1]). In particular the torsion-free component of Ω_*^G is all of even degree.

Apart from the non-abelian groups which are metacyclic, there is no other finite non-abelian group on which the evenness conjecture has been shown to hold.

The main difficulty lies in the understanding of the equivariant bordism groups $\Omega_*^G\{\mathcal{F}\}(^{\widetilde{G}}B_G U($ of the classifying spaces $^{\widetilde{G}}B_G U(n)$ associated to $\mathbb{S}^1$-central extensions $\widetilde{G}$ of G for different families $\mathcal{F}$ of subgroups. These bordism groups are the ones appearing once we try to calculate the equivariant unitary bordism groups for adjacent pair of families. Any development on the understanding of these equivariant unitary bordism groups will shed a light on the proof of the evenness conjecture for a bigger class of groups.

5 Conclusion

The evenness conjecture for equivariant unitary bordism has been an important question in algebraic topology for more than forty years. The conjecture has been proved to hold only for compact abelian Lie groups and finite metacyclic groups, for all other groups the conjecture remains open. We do hope that the present summary of known results will help settle the conjecture in the near future.

References

A. Ángel, J. M. Gómez, and B. Uribe (n.d.). “Equivariant complex bundles, fixed points and equivariant unitary bordism”. arXiv: 1710.00879 (cit. on pp. 1236, 1239, 1241, 1242, 1244, 1256).

M. F. Atiyah (1961). “Bordism and cobordism”. *Proc. Cambridge Philos. Soc.* 57, pp. 200–208. MR: 0126856 (cit. on p. 1245).

M. F. Atiyah and F. Hirzebruch (1961). “Vector bundles and homogeneous spaces”. In: *Proc. Sympos. Pure Math., Vol. III*. American Mathematical Society, Providence, R.I., pp. 7–38. MR: 0139181 (cit. on p. 1245).

Michael Bix and Tammo tom Dieck (1978). “Characteristic numbers of G-manifolds and multiplicative induction”. *Trans. Amer. Math. Soc.* 235, pp. 331–343. MR: 0478182 (cit. on p. 1241).

P. E. Conner and E. E. Floyd (1966). “Maps of odd period”. *Ann. of Math. (2)* 84, pp. 132–156. MR: 0203738 (cit. on pp. 1237, 1239).

P. E. Conner and Larry Smith (1969). “On the complex bordism of finite complexes”. *Inst. Hautes Études Sci. Publ. Math.* 37, pp. 117–221. MR: 0267571 (cit. on p. 1246).

Tammo tom Dieck (1972). “Orbittypen und äquivariante Homologie. I”. *Arch. Math. (Basel)* 23, pp. 307–317. MR: 0310919 (cit. on pp. 1237, 1238).

– (1973). “Equivariant homology and Mackey functors”. *Math. Ann.* 206, pp. 67–78. MR: 0375285 (cit. on p. 1257).

– (1987). *Transformation groups*. Vol. 8. De Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin, pp. x+312. MR: 889050 (cit. on p. 1237).

S. O. Kochman (1996). *Bordism, stable homotopy and Adams spectral sequences*. Vol. 7. Fields Institute Monographs. American Math. Society, Providence, RI, pp. xiv+272. MR: 1407034 (cit. on pp. 1245, 1246).

Peter S. Landweber (1971). "Complex bordism of classifying spaces". *Proc. Amer. Math. Soc.* 27, pp. 175–179. MR: 0268885 (cit. on p. 1246).

– (1972). "Equivariant bordism and cyclic groups". *Proc. Amer. Math. Soc.* 31, pp. 564–570. MR: 0296969 (cit. on pp. 1236, 1245, 1247, 1249).

Connor Lazarov (1972). "Actions of groups of order pq". *Trans. Amer. Math. Soc.* 173, pp. 215–230. MR: 0309131 (cit. on pp. 1236, 1256).

Peter Löffler (1974). "Bordismengruppen unitärer Torusmannigfaltigkeiten". *Manuscripta Math.* 12, pp. 307–327. MR: 0348776 (cit. on pp. 1236, 1245).

Wolfgang Lück and Bernardo Uribe (2014). "Equivariant principal bundles and their classifying spaces". *Algebr. Geom. Topol.* 14.4, pp. 1925–1995. MR: 3331607 (cit. on p. 1250).

J. P. May (1996). *Equivariant homotopy and cohomology theory*. Vol. 91. CBMS Regional Conference Series in Mathematics. With contributors M. Cole, G. Comezaña, S. Costenoble, A. D. Elmendorf, J. P. C. Greenlees, L. G. Lewis, Jr., R. J. Piacenza, G. Triantafillou, and S. Waner. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, pp. xiv+366. MR: 1413302 (cit. on pp. 1235, 1236, 1245).

J. Milnor (1960). "On the cobordism ring Ω^* and a complex analogue. I". *Amer. J. Math.* 82, pp. 505–521. MR: 0119209 (cit. on pp. 1236, 1245).

S. P. Novikov (1960). "Some problems in the topology of manifolds connected with the theory of Thom spaces". *Soviet Math. Dokl.* 1, pp. 717–720. MR: 0121815 (cit. on pp. 1236, 1245).

Erich Ossa (1972). "Unitary bordism of abelian groups". *Proc. Amer. Math. Soc.* 33, pp. 568–571. MR: 0293666 (cit. on pp. 1236, 1245, 1247, 1255).

Russell J. Rowlett (1978). "The fixed-point construction in equivariant bordism". *Trans. Amer. Math. Soc.* 246, pp. 473–481. MR: 515553 (cit. on p. 1257).

– (1980). "Bordism of metacyclic group actions". *Michigan Math. J.* 27.2, pp. 223–233. MR: 568643 (cit. on pp. 1236, 1256).

R. E. Stong (1970). "Complex and oriented equivariant bordism". In: *Topology of Manifolds (Proc. Inst., Univ. of Georgia, Athens, Ga., 1969)*. Markham, Chicago, Ill., pp. 291–316. MR: 0273644 (cit. on pp. 1236–1238, 1245, 1247, 1250–1252).

René Thom (1954). "Quelques propriétés globales des variétés différentiables". *Comment. Math. Helv.* 28, pp. 17–86. MR: 0061823 (cit. on pp. 1236, 1245).

Received 2018-02-19.

Bernardo Uribe
Departamento de Matemáticas y Estadística
Universidad del Norte
Km. 5 via Puerto Colombia, Barranquilla
Colombia
bjongbloed@uninorte.edu.co

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1259–1280)

LITTLE DISKS OPERADS AND FEYNMAN DIAGRAMS

Thomas Willwacher

Abstract

The little disks operads are classical objects in algebraic topology which have seen a wide range of applications in the past. For example they appear prominently in the Goodwillie-Weiss embedding calculus, which is a program to understand embedding spaces through algebraic properties of the little disks operads, and their action on the spaces of configurations of points (or disks) on manifolds. In this talk we review the recent understanding of the rational homotopy theory of the little disks operads, and how the resulting knowledge can be used to fulfil the promise of the Goodwillie-Weiss calculus, at least in the "simple" setting of long knot spaces and over the rationals. The derivations prominently use and are connected to graph complexes, introduced by Kontsevich and other authors.

1 Introduction

The little disks operads are collections of spaces $\mathrm{D}_n(r)$ of rectilinear embeddings of r little disks in the n-dimensional unit disk

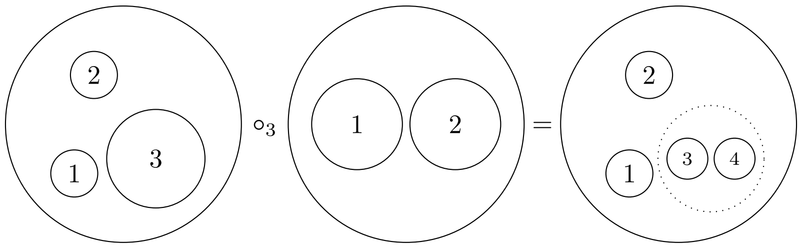
$$\mathrm{D}_n(r) = \mathrm{Emb}^{rl}(\mathbb{D}_n^{\sqcup r}, \mathbb{D}_n).$$

Here rectilinear means that the embedding may rescale and translate the little disks, but not rotate or otherwise deform them. The operadic compositions are defined through the gluing of configurations of disks, with one configuration being

The author acknowledges partial support by the European Research Council, ERC StG 678156–GRAPHCPX..

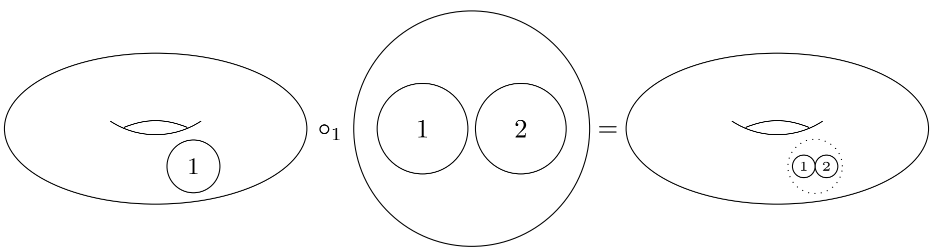
MSC2010: primary 18D50; secondary 55P62, 57Q45, 81T18, 57R56.

inserted in place of small disk as the following example illustrates.

(1) 

The framed little n-disks operad D_n^{fr} is a variant in which one allows the embeddings to also rotate the little disks.

The little disks operads have seen various applications in algebra, topology, and even some branches of mathematical physics. We will focus on one particular relatively recent application here, the manifold calculus of Weiss and Goodwillie [1999] and Weiss [1999]. To this end let M be a manifold of dimension m. We may consider the spaces of embeddings of m-dimensional disks in M, $\mathrm{conf}_M(r) = \mathrm{Emb}(\mathbb{D}_m^r, M)$. Again by composition of embeddings (i.e., gluing of disks) the operad D_m (and likewise D_m^{fr}) naturally acts on the the collection of spaces conf_M.



The idea of the Goodwillie-Weiss manifold calculus is then that properties of the space M, and spaces derived from it, may be accessed using the spaces $\mathrm{conf}_M(r)$ and the action of $\mathrm{D}_n^{(fr)}$ upon them. In particular, for N another manifold of dimension n, and under the technical condition that $n \geq m+3$, one can express the space of embeddings from M into N as a derived mapping space between the right D_m^{fr}-modules conf_M and conf_N, see the work of Weiss and Boavida de Brito [2013],

$$\mathrm{Emb}(M, N) \simeq \mathrm{Map}^h_{mod-\mathrm{D}_m^{fr}}(\mathrm{conf}_M, \mathrm{conf}_N). \tag{2}$$

In other words, the manifold calculus replaces the complicated topological space of knottings of M in N on the left by a (potentially) accessible algebraic object on the right. One problem of this approach had been that the algebraic object on the right is still relatively complicated and hard to understand. However, due to recent progress in understanding the rational (or real) homotopy theory of the little disks operads and configuration spaces Campos and Willwacher [2016] and Idrissi [2016], information about the right-hand side may be obtained.

For this exposition we will in fact restrict to the simplest setting, when $M = \mathbb{R}^m$ and $N = \mathbb{R}^n$. In this case one studies the spaces of long knots $\mathrm{Emb}_\partial(\mathbb{R}^m, \mathbb{R}^n)$, i.e., of embeddings of $\mathbb{R}^m$ into $\mathbb{R}^n$ which agree with the standard embedding outside of a compact. For technical reasons one furthermore reduces to the homotopy fiber over immersions

$$\overline{\mathrm{Emb}}_\partial(\mathbb{R}^m, \mathbb{R}^n) = \mathrm{hofiber}\left(\mathrm{Emb}_\partial(\mathbb{R}^m, \mathbb{R}^n) \to \mathrm{Imm}_\partial(\mathbb{R}^m, \mathbb{R}^n)\right).$$

The appropriate version of the embedding calculus for this situation then states that for $n - m \geq 3$ there is a weak equivalence (cf. Weiss and Boavida de Brito [2015], Ducoulombier and Turchin [2017], and Dwyer and Hess [2012])

$$\overline{\mathrm{Emb}}_\partial(\mathbb{R}^m, \mathbb{R}^n) \simeq \Omega^{m+1}\mathrm{Map}^h_{op}(\mathrm{D}_m, \mathrm{D}_n), \tag{3}$$

where on the right-hand side we have the $m+1$-fold loop space of the derived mapping space of operads between D_m and D_n. In particular, note that the Goodwillie-Weiss calculus states that the homotopy type of the space of (codimension $\geq$ 3-)knots is already fully encoded in the homological algebra of the relatively "simple" insertion operations (1).

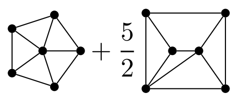
The final result we want to study is that the rational homotopy type of the right-hand side of (3) may be computed and expressed fully in terms of combinatorial data, through graph complexes and algebraic structures on graph complexes. The results we review here are mostly taken from joint works of the author Fresse and Willwacher [2015] and Fresse, Turchin, and Willwacher [2017b]. We also refer to these works for more technical details, which often have to be omitted from our exposition for reasons of brevity.

Notation and conventions. We generally work over the ground field $\mathbb{Q}$ unless otherwise stated, i.e., all vector spaces, commutative algebras etc. are considered over $\mathbb{Q}$. As usual we abbreviate the phrase "differential graded" by dg, and "differential graded commutative algebra" by dgca. We omit the prefix dg if clear from the context. For example "vector space" will typically mean dg vector space. We generally work in cohomological conventions, so that all of our differentials have degree +1. For a (dg) vector space V we denote by $V[k]$ the degree shifted vecor space. If $v \in V$ has degree d then the corresponding object in $V[k]$ has degree

$d - k$. For an introduction to the language of operads we refer the reader to the textbooks Loday and Vallette [2012] or Fresse [n.d.], whose notation we shall essentially follow. A standard reference for the little disks operads is May [1972, section 4].

2 Graph complexes

Graph complexes are differential graded vector spaces of linear combinations or series of combinatorial graphs. There are various versions depending on the type of graphs considered, for example complexes of undirected graphs, directed acyclic graphs, ribbon graphs etc. Here we will consider only the simplest version, as introduced by M. Kontsevich. We define GC_n to be the $\mathbb{Q}$-vector space of series of isomorphism classes of admissible graphs. Here an admissible graph is a connected undirected graph with an orientation, all of whose vertices have valence ≥ 2, and which does not have odd symmetries, a condition we shall elucidate shortly.

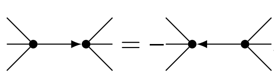


The definition depends on an integer n, which determines the cohomological degree of graphs, with a graph in GC_n being assigned degree

$$(\#\text{vertices} - 1)n - (\#\text{edges})(n - 1).$$

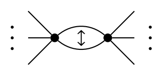
In other words, we consider the vertices as carrying degree n, and the edges as carrying degree $1-n$. An orientation on a graph Γ is the following data, depending on the parity of n:

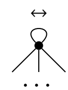
- For n even, an orientation of Γ is an ordering or of the set of edges of Γ. If two such orderings or, or′ differ by a permutation σ, we identify the oriented graphs up to sign
$$(\Gamma, \mathrm{or}) = \mathrm{sgn}(\sigma)(\Gamma, \mathrm{or}').$$
- For n odd an orientation consists of an ordering of the set of vertices and half-edges. Again we identify two such orderings up to sign. Note that providing an ordering of the set of half-edges up to sign is equivalent to providing a direction on edges, identifying directions up to sign



The presence of the orientation implies that graphs with orientation reversing (odd) symmetries yield zero vectors in the graph complex. More concretely, for n even and odd symmetry of a graph is an automorphism inducing an odd permutation on the set of edges. Similarly, for n odd an odd symmetry is an automorphism inducing an odd permutation on the set of half-edges and vertices.

Note that this in particular implies that for n even any graph with a double edge is considered zero, for it has an odd symmetry by swapping the edges in the double edge. Likewise, for n odd any graph with a tadpole (or short cycle) is zero due to the symmetry reversing the cycle.

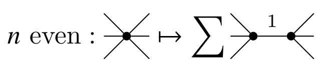


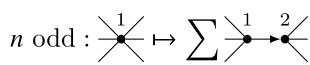


We define on GC_n a differential d, splitting vertices of graphs. More concretely

$$d\Gamma = \sum_v split(\Gamma, v)$$

with the operation $split(\Gamma, v)$ replacing the vertex v by two vertices connected by an edge and summing over all ways of reconnecting the edges incident at v to the two new vertices. Here the orientation on graphs in $split(\Gamma, v)$ is chosen so that for n even the newly created edge becomes the first in the ordering of edges. For n odd assume without loss of generality that v is the first vertex in the ordering. The orientation is chosen such that the newly created vertices are the first two in the ordering, with the newly created edge pointing from the first to the second.





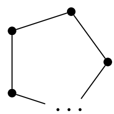
It is an easy exercise to check that with these conventions on the orientation $d^2 = 0$, so that we can consider the graph cohomology $H(\mathrm{GC}_n) = \ker d/\mathrm{im} d$. This cohomology is a somewhat mysterious object that can at present only partially be computed. Let us recall a few known facts.

First note that the differential cannot alter the loop order of a graph, and hence the graph complex decomposes into a direct product of subcomplexes of fixed loop order $\mathrm{GC}_n^{k\text{-loop}}$. Furthermore, GC_n depends on n essentially only up to parity, and hence one can see that

$$H^j(\mathrm{GC}_{n+2}^{k\text{-loop}}) \cong H^{j+2k}(\mathrm{GC}_n^{k\text{-loop}}).$$

In particular, knowing $H(\mathrm{GC}_n)$ for one even and one odd n suffices to determine $H(\mathrm{GC}_n)$ for all n. On the other hand, despite considerable effort, the author has not found any relation between $H(\mathrm{GC}_n)$ and $H(\mathrm{GC}_{n+1})$ that would allow for the computation of one through the other.

In low loop orders the graph cohomology can be computed explicitly by hand or with the help of computers. For example, in loop order 1 the most general graph has the form

(4) $L_k =$  (k vertices and k edges, deg$=n-k$).

These graphs have odd symmetries, and hence vanish as elements of the graph complex, unless $k \equiv 2n+1 \mod 4$, so that

$$H(\mathrm{GC}_n^{1\text{-loop}}) \cong \prod_{\substack{k\geq 1 \\ k\equiv 2n+1 \mod 4}} \mathbb{Q}W_k, \tag{5}$$

with the class W_k living in cohomological degree $k-n$.

For loop order $k \geq 2$ one can show that the inclusion of the subcomplex

$$\mathrm{GC}_n^{k\text{-loop},\geq 3} \subset \mathrm{GC}_n^{k\text{-loop}}$$

spanned by graphs all of whose vertices are at least trivalent is a quasi-isomorphism, cf. Kontsevich [1993] and Willwacher [2015b]. In this subcomplex, a graph with v vertices must then have at least $\frac{3}{2}v$ edges, and hence one can derive the simple upper degree bound for $k \geq 2$

$$H^{>-k(n-3)-3}(\mathrm{GC}_n^{k\text{-loop}}) = 0. \tag{6}$$

By somewhat different methods one can also derive a lower degree bound (cf. Willwacher [2015b])

$$H^{<-k(n-2)-1}(\mathrm{GC}_n^{k\text{-loop}}) = 0.$$

Next, we shall note that GC_n carries the structure of a dg Lie algebra. The Lie bracket is defined combinatorially by inserting a graph in vertices of another.

$$[\gamma, \nu] = \gamma \bullet \nu - (-1)^{|\gamma||\nu|}\nu \bullet \gamma$$

with

$$\gamma \bullet \nu := \sum_{x \in V\gamma} \gamma(\text{insert } \nu \text{ in place of } x)$$

In the case of $n = 2$ the author showed the following.

Theorem 1 (Willwacher [ibid.]). The zeroth cohomology $H^0(\mathrm{GC}_2)$ can be identified with the (completed) Grothendieck-Teichmüller Lie algebra $\mathfrak{grt}_1$. Furthermore $H^1(\mathrm{GC}_2) \cong \mathbb{K}$ and $H^{\leq 1}(\mathrm{GC}_2) = 0$.

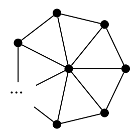
We will not recall the somewhat technical definition of the Grothendieck-Teichmüller Lie algebra $\mathfrak{grt}_1$ defined by Drinfeld [1990], but rather recall the following important result of Francis Brown.

Theorem 2 (Brown [2012]). There is an injective Lie algebra morphism

$$F_{Lie}(\sigma_3, \sigma_5, \sigma_7, \dots) \to \mathfrak{grt}_1$$

from the complete free Lie algebra in generators $\sigma_3, \sigma_5, \dots$.

Both results together yield an infinite family of nontrivial graph cohomology classes. One can provide explicit integral formulas for the graph cocycles representing σ_{2j+1} as in "P. Etingof's conjecture about Drinfeld associators" [2014]. Concretely, σ_{2k+1} is represented by a linear combination of diagrams of loop order $2k+1$

$\sigma_{2k+1} \simeq$  $+ (\cdots),$

where the first "wheel" graph has $2k+2$ vertices, and the terms (...) on the right which are not explicitly written are linear combinations of graphs all of whose vertices have valence $\leq 2k$.

We shall not recall here in detail the known results for the graph cohomology $H(\mathrm{GC}_3)$, or equivalently $H(\mathrm{GC}_n)$ for n odd. Let us just mention that a large family of non-trivial cohomology classes in top degree is known through Chern-Simons theory. Furthermore there are conjectures regarding the precise shape of the top degree cohomology, see Vogel [2011] and Kneissler [2000, 2001a,b] for details.

Computer generated tables of the numbers $\dim H^j(\mathrm{GC}_n^{k-\mathrm{loop}})$ can be found in Figure 1. The red lines depict the degree bounds beyond which the cohomology is zero. The cohomology classes giving rise to the numbers appearing in the (computer accessible portion of) the tables can be "explained", see Khoroshkin, Willwacher, and Živković [2017] for more details. However, at large we still do not know what $H(\mathrm{GC}_n)$ is, and in particular which entries of the table are zero. To this end, let us just note a famous vanishing conjecture which goes back to Kontsevich, and in a similar form to Drinfeld.

Conjecture 3. $H^1(\mathrm{GC}_2) = 0$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
7	1	0	0	0	0	0	0	0	0	1				
6	0	0	0	0	0	0	0	0	0	0				
5	0	0	0	0	0	0	0	0	0	0				
4	0	0	0	0	0	0	0	0	0	0				
3	1	0	0	0	0	1	0	1	1	2				
2	0	0	0	0	0	0	0	0	0	0				
1	0	0	0	0	0	0	0	0	0	0	0			
0	0	0	1	0	1	0	1	1	1	1	2	2	3	
-1	1	0	0	0	0	0	0	0	0	0	0	0	0	

	1	2	3	4	5	6	7	8	9	10	11	12
8	1											
4	1											
0	1											
-2	0	0	0	0	0	0	0	0	0	0	0	0
-3	0	1	1	1	2	2	3	4	5	6	8	9
-4	0	0	0	0	0	0	0	0				
-5	0	0	0	0	0	0	0	0				
-6	0	0	0	0	0	1	1	2				
-7	0	0	0	0	0	0	0	0				
-8	0	0	0	0	0	0	0	0	0			

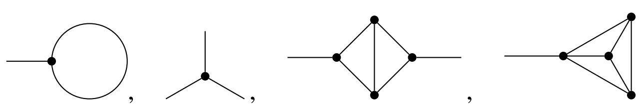
Figure 1: Computer generated tables of $\dim H^j(\mathrm{GC}_2^{k\text{–loop}})$ (top) and $\dim H^j(\mathrm{GC}_3^{k\text{–loop}})$ (bottom), with j being the row index and the loop order k the column index. All entries above the upper and below the lower red line are zero. Entries between the red lines which are not shown have not been computed.

	Even	Odd		Even	Odd
loop order	$\tilde{\chi}_b^{even}$	$\tilde{\chi}_b^{odd}$	loop order	$\tilde{\chi}_b^{even}$	$\tilde{\chi}_b^{odd}$
1	0	1	16	-3	6
2	1	1	17	-1	4
3	0	1	18	8	-5
4	1	2	19	12	-14
5	-1	1	20	27	-21
6	1	2	21	14	-11
7	0	2	22	-25	21
8	0	2	23	-39	44
9	-2	1	24	-496	504
10	1	3	25	-2979	2969
11	0	1	26	-412	413
12	0	3	27	38725	-38717
13	-2	4	28	10583	-10578
14	0	2	29	-667610	667596
15	-4	2	30	28305	-28290

Figure 2: Table of the Euler characteristics of the graph complexes GC_n for even and odd n from Willwacher and Živković [2015]. Note that for high loop orders the Euler characteristics for the even and odd complexes are astonishingly similar, with the sign difference being due to conventions.

As a final remark we shall mention that while the above tables suggest that the cohomology of the graph complexes $H(\mathrm{GC}_n)$ for even and odd n is very different. However, the Euler characteristic computations of GC_n from Willwacher and Živković [2015], which we reproduce in Figure 2, show that at least the Euler characteristics of both complexes in high loop orders are strikingly similar. The author can currently not explain this fact.

2.1 A variant with external legs. We will also need a slight variant of the above graph complexes. We may consider a complex of graphs $\mathrm{HGC}_{m,n}$ built using graphs with "external legs" or hairs, as shown in the following pictures



We require that all non-hair vertices are at least trivalent. The cohomological degree of a graph is determined by the formula

$$n(\#\text{internal vertices}) - (n-1)(\#\text{edges}) + m(\#\text{internal vertices} - 1).$$

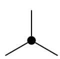
Here we count the edges being part of a hair as edges as well. One also equips these hairy graphs with an orientation. Graphs that possess odd symmetries are hence considered zero in the graph complex. The differential is again given by splitting (non-hair) vertices. This operation cannot change the number of hairs, nor the first Betti number (i.e„ the number of loops) of a graph, Hence the complex $\mathrm{HGC}_{m,n}$ splits into a direct product of subcomplexes of fixed number of hairs and loops

$$\mathrm{HGC}_{m,n}^{k\text{-loop},h\text{-hair}} \subset \mathrm{HGC}_{m,n}.$$

In general, the cohomology of these complexes is not known, but at least one has partial information. In low loop orders, or for low numbers of hairs one can explicitly compute the cohomology. For example, in loop order zero the cohomology is precisely one dimensional, and represented by the following graph cocycles:

(7)

 for $n-m$ even, in cohomological degree $m-n+1$

 for $n-m$ odd, in cohomological degree $2(m-n)+3$

The computation of the loop order one cohomology we leave to the reader. The computation in loop order two can be found in Conant, Costello, Turchin, and Weed [2014].

Next, using that non-hair vertices can be required to be at least trivalent one can easily derive the upper degree bounds

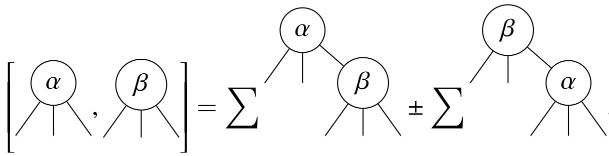
$$H^{>-k(n-3)-(h-1)(n-m-2)-1}(\mathrm{HGC}_{m,n}^{k\text{-loop},h\text{-hair}}) = 0. \tag{8}$$

For example, we will use below that for $n-3 \geq m$ the only non-trivial classes in degree $-m$ live in loop order $k=0$ and are hence found in the list (7).

By other methods (see Arone and Turchin [2014] and Willwacher [2015b]) one can also obtain the lower degree bound

$$H^{<-k(n-2)-(h-1)(n-m-1)}(\mathrm{HGC}_{m,n}^{k\text{-loop},h\text{-hair}}) = 0.$$

Furthermore, the complexes $\mathrm{HGC}_{m,n}$ are also dg Lie algebras. The Lie bracket of two hairy graphs is computed by connecting a hair of one graph to vertices of the other as indicated in the following picture:



For further information about the hairy graph cohomology $H(\mathrm{HGC}_{m,n})$, we refer the reader to Arone and Turchin [2015] containing a computation of the Euler characteristic, to Khoroshkin, Willwacher, and Živković [2015] containing a construction of infinite series of nontrivial classes and numerical results, or to Fresse, Turchin, and Willwacher [2017b] for general information.

3 The little disks operads

3.1 Cohomology of the little disks operad. The cohomology of the little disks operads D_n has been computed by Arnold (for $n = 2$) and F. Cohen (for all n).

Theorem 4 (Arnold [1969] and Cohen [1976]). For $n \geq 2$ and $r \geq 1$ the cohomology algebra of the space $\mathrm{D}_n(r)$ has the presentation

$$H(\mathrm{D}_n(r)) = \mathbb{Q}[\omega_{ij} \mid 1 \leq i \neq j \leq n]/\langle R\rangle,$$

where ω_{ij} are generators of degree $n-1$ and the relations R read

$$\omega_{ij} = (-1)^n \omega_{ji} \qquad \omega_{ij}^2 = 0 \qquad \omega_{ij}\omega_{jk} + \omega_{jk}\omega_{ki} + \omega_{ki}\omega_{ij} = 0$$

From the operad structure on D_n the collection of graded commutative algebras $H(\mathrm{D}_n)$ receives a cooperad structure. Generally, we will call a cooperad in the category of dg commutative algebras a Hopf cooperad, so that in particular $H(\mathrm{D}_n)$ is a Hopf cooperad.

To understand the operad structure it is slightly easier to consider the dual operad $H_\bullet(\mathrm{D}_n)$. For $n = 1$ this is just the associative operad. For $n \geq 2$ it can be identified with the n-Poisson operad Pois_n, which generated by a commutative product operation $\wedge$ of degree 0 and a Lie bracket $[,]$ of degree $1-n$ satisfying the compatibility relation

$$[x_1, x_2 \wedge x_3] = [x_1, x_2] \wedge x_3 + x_2 \wedge [x_1, x_3].$$

3.2 Rational homotopy theory of operads. Rational homotopy theory is the study of rational homotopy types of spaces. In Sullivan's approach, the main ingredient is a Quillen adjunction between the categories of simplicial sets (which we shall think of as topological spaces) and of dg commutative algebras

$$\Omega : \mathrm{sSet} \leftrightarrows \mathrm{dgca}^{op} : G.$$

Sullivan's functor Ω sends a simplicial set X to the dg commutative algebra of piecewise polynomial differential forms

$$\Omega(X) := \mathrm{Hom}_{\mathrm{sSet}}(X, \Omega_{poly}(\Delta^\bullet)).$$

We readily extend the definition to topological spaces instead of simplicial sets via the singular simplicial complexes functor $\mathrm{Sing}_\bullet$. We shall quietly abuse the notation and write, for a topological space X,

$$\Omega(X) := \Omega(\mathrm{Sing}_\bullet X).$$

When X is a manifold $\Omega(X) \otimes_{\mathbb{Q}} \mathbb{R}$ is weakly equivalent to the dg commutative algebra of de Rham differential forms on X. For our purposes, rational homotopy theory can be seen as the study of the quasi-isomorphism type of the dg commutative algebra $\Omega(X)$.

Let us next consider a topological operad $\mathcal{T}$. To study its rational homotopy type we would like to apply the functor Ω and study the resulting cooperad object in dg commutative algebras. Unfortunately, due to incompatible monoidality properties of the functor Ω the collection $\Omega(\mathcal{T})$ is not naturally a cooperad. More concretely, the problem here is that one has a natural quasi-isomorphism $\Omega(X) \otimes \Omega(Y) \to \Omega(X \times Y)$, but no natural morphism in the other direction. There are essentially three known approaches to work around this technical problem, by (i) using operads up to homotopy or (ii) changing the functor Ω or (iii) to use completed tensor products in the smooth setting.

While all three approaches have been used in the literature, we use here approach (ii). We shall follow Fresse's rational homotopy theory for operads, see Fresse [n.d.] and Fresse and Willwacher [2015, section 0], which we briefly outline. For brevity we call a cooperad in dg commutative algebras a Hopf cooperad. Fresse constructs a Quillen adjunction

$$\Omega_\sharp : \mathrm{sSet\text{-}Op} \leftrightarrows \mathrm{Hopf\text{-}Op}^c : G$$

between the model categories of reduced operads in simplicial sets, and that of dg Hopf cooperads. The functor $\Omega_\sharp$ here is defined as left adjoint of the realization functor G and shall be seen as an operadic upgrade of Sullivan's functor Ω. In each arity, $\Omega_\sharp$ is weakly equivalent to Ω.

For our purposes, studying the rational homotopy type of a topological operad $\mathcal{T}$ amounts to studying the quasi-isomorphism class of the dg Hopf cooperad $\Omega_\sharp(\mathcal{T})$, where we again quietly extend $\Omega_\sharp$ to topological spaces instead of simplicial sets, taking singular simplices. Furthermore, if $\mathcal{S}$ and $\mathcal{T}$ are simplicial (or topological) operads, then one can use the Quillen adjunction above to compute

$$\mathrm{Map}_{\text{sSet-Op}}(\mathcal{S}, \mathcal{T}^{\mathbb{Q}}) = \mathrm{Map}_{\text{sSet-Op}}(\mathcal{S}, G(\Omega_\sharp(\mathcal{T}))) \simeq \mathrm{Map}_{HopfOp^c}(\Omega_\sharp(\mathcal{T}), \Omega_\sharp(\mathcal{S})).$$

We finally note that Fresse's framework has the downside that it requires our operads to be reduced, i.e., that $\mathcal{T}(0) = \mathcal{T}(1) = *$ is a point. Unfortunately the little disks operads introduced above are not reduced, since $\mathrm{D}_n(1)$ is not a point, only contractible. However, there are homotopy equivalent variants of D_n, for example the Fulton-MacPherson operad FM_n Getzler and Jones [1994], which are reduced. Generally, an E_n operad is a topological operad weakly equivalent to D_n. In the following we will abuse the notation a bit and denote by E_n some chosen reduced operad weakly equivalent to D_n. Furthermore, for technical reasons the arity zero operations in $\mathcal{T}(0) = *$ are encoded in a Λ-structure instead of considering them as operations in the operad. A Λ-structure is the collection of all possible composition maps with nullary operations $\mathcal{T}(r+s) \to \mathcal{T}(r)$, which are required to satisfy natural compatibility relations. For simplicity of notation we will hide this further technical complication and do not mark the presence of the Λ-structure in our notation. We refer to Fresse's book Fresse [n.d.] for details. It is shown in Fresse, Turchin, and Willwacher [2017a] that the mapping spaces computed in the category of reduced operads are weakly equivalent to those computed in the full category of operads, thus justifying our restriction to the reduced setting.

3.3 Formality and intrinsic formality of E_n operads. Today the rational homotopy types of the little disks operads are fully understood through the following Theorem.

Theorem 5 (Formality Theorem for the E_n operads). The dg Hopf cooperads $\Omega_\sharp(E_n)$ are formal, i.e., connected by a chain of weak equivalences to the cohomology cooperad $\mathrm{e}_n^c = H(E_n)$.

The Theorem has the following history. The formality of E_2 was first shown by Tamarkin [2003]. The statement for higher n was first established by Kontsevich [1999], albeit over the ground ring $\mathbb{R}$. It has then been noted in Guillén Santos, Navarro, Pascual, and Roig [2005] that Kontsevich's statement can be improved to yield formality over $\mathbb{Q}$, provided one disregards the arity zero operations in the cooperads. Finally, the remaining statement of formality over $\mathbb{Q}$ with zero-ary operations was shown by Fresse and the author in Fresse and Willwacher [2015]. Also, surprisingly, we were able to show a significantly stronger result. A dg Hopf cooperad $\mathfrak{C}$ is called intrinsically formal if any dg Hopf cooperad $\mathfrak{D}$ with $H(\mathfrak{D}) \simeq H(\mathfrak{C})$ is weakly equivalent to $\mathfrak{C}$. One then has:

Theorem 6 (Intrinsic formality for the little diks operadsFresse and Willwacher [ibid.]). The dg Hopf cooperad $\Omega_\sharp(E_n)$ is intrinsically formal for $n \geq 3$ and n not divisible by 4. If $n \geq 3$ is divisible by 4, one does not have intrinsic formality, but the following statement is retained: Suppose that $\mathfrak{D}$ is another dg Hopf cooperad such that $H(D) \cong H(E_n)$. Suppose further that $I : \mathfrak{D} \to \mathfrak{D}$ is an involution which agrees on cohomology with the canonical involution of E_n by mirror reflection along a coordinate axis. Then $\mathfrak{D} \simeq H(E_n)$.

For $n = 2$ the analogous statement is an open conjecture.

Conjecture 7. The little 2-disks operad is rationally intrinsically formal.

In fact, this conjecture would follow from the vanishing Conjecture 3 on the graph cohomology above.

To conclude this subsection let us also remark on several closely connected formality questions for the little disks operads. First we note that the little disks operads come with natural maps $\mathrm{D}_m \to \mathrm{D}_n$. One may ask what the rational (or real) homotopy type of these maps are. The question has been answered in works of Turchin and the author Turchin and Willwacher [2014] and Fresse and the author Fresse and Willwacher [2015], improving upon earlier results by Lambrechts and Volić Loday and Vallette [2012].

Theorem 8 (Fresse and Willwacher [2015], Turchin and Willwacher [2014], and Loday and Vallette [2012]). Let $n \geq m \geq 1$. Then the map $E_m \to E_n$ is rationally formal for $n - m \neq 1$, and not formal (even over $\mathbb{R}$) for $n - m = 1$.

Furthermore the real homotopy type of the map $E_{n-1} \to E_n$ can be fully described.

In a different direction, note that the group $O(n)$ naturally acts on the little n-disks operad $\mathrm{D}(n)$. One can hence ask whether the operad $\mathrm{D}(n)$ is $O(n)$-equivariantly formal or not. This is equivalent to asking whether the framed little disks operads are formal or not. This latter question has also been answered by now: The framed little 2-disks operad D_2^{fr} was shown to be formal by Giansiracusa and Salvatore [2010] (over $\mathbb{R}$) and independently by Ševera [2010] (over $\mathbb{Q}$).

Furthermore, the formality (over $\mathbb{R}$) of D_n^{fr} for n is even is shown in Khoroshkin and Willwacher [2017]. For $n \geq 3$ odd the operad D_n^{fr} is not formal, but explicit "small" models capturing the real homotopy type can be found Khoroshkin and Willwacher [2017] and Moriya [2017]

At present, the (arguably) most important remaining open problem regarding the formality of the little disks operads is the following.

Open Problem 9. Determine the $O(m) \times O(n-m)$-equivariant rational homotopy type of the Map $D_m \to D_n$.

This homotopy type appears for example in connection to the Goodwillie-Weiss manifold calculus, see Section 5 below.

3.4 Kontsevich's graphical model for E_n. The intrinsic formality statement (Theorem 6 above) can be shown by an analysis of the graph complex GC_n discussed in Section 2. The object that builds the bridge are the graphical models Graphs_n for the E_n opeads introduced by Kontsevich [1999]. More concretely, one defines a collection of dg commutative algebras $\mathrm{Graphs}_n(r)$ as follows. The space $\mathrm{Graphs}_n(r)$ is the space of linear combinations of isomorphism classes of graphs of the following type:

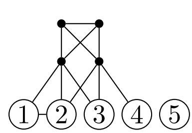
- The graph is an undirected graph with r numbered "external" vertices, and an arbitrary (but finite) number of "internal" vertices.

- All internal vertices have valence at least 2.

- Every connected component contains at least one external vertex.

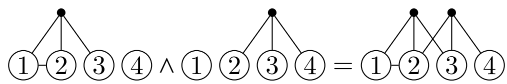
- Graphs are equipped with an orientation, as in Section 2, and we identify orientations up to sign. More concretely, for n even an orientation is an ordering of the edges, while for n odd an orientation is an odering of the set of half-edges and vertices. As discussed before, the presence of the orientation renders graphs with odd symmetries zero

The following is an examples of a graph in $\mathrm{Graphs}_n(5)$.



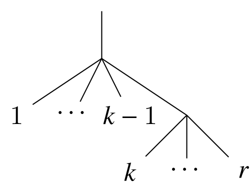
The space $\mathrm{Graphs}_n(r)$ is equipped with a differential by edge contraction, schematically:

Furthermore, we have a commutative product by gluing two graphs along the external vertices.



It is clear that $\mathrm{Graphs}_n(r)$ is free as a graded commutative algebra, generated by graphs that are internally connected, i.e., connected after removal of the external vertices.

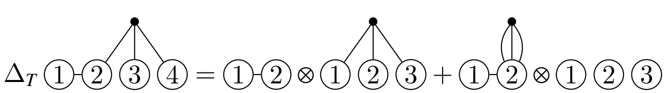
Furthermore there is cooperad structure on the collection Graphs_n, with the coproduct being the contraction of subgraphs. More precisely, for the generating cocomposition Δ_T corresponding to a tree



we have that

$$\Delta_T(\Gamma) = \sum_{\gamma} \Gamma/\gamma \otimes \gamma, \tag{9}$$

where the sum is over all subgraphs $\gamma \subset \Gamma$ containing the external vertices $k, \dots, r$ and no other external vertices, and the graph Γ/γ is obtained by contracting γ to one external vertex. Here is an example for $r = 4$ and $k = 2$:



In this formulas, and more generally (9) there is a natural way to define an orientation on the graphs on the right-hand side of the equation, given an orientation on the left-hand side, thus fixing the signs.

Finally, we note that for $n \geq 2$ we have a map of dg Hopf cooperads

$$\mathrm{Graphs}_n \to H(\mathrm{D}_n) = \mathrm{e}_n^c,$$

which is defined by sending any graph with internal vertices to zero, and sending a graph Γ without internal vertices and edges $(i_1, j_1), \dots, (i_k, j_k)$ to

$$\omega_{i_1 j_1} \cdots \omega_{i_k j_k},$$

cf. Theorem 4. Kontsevich and Lambrechts-Volić have then shown:

Proposition 10 (Kontsevich [1999] and Loday and Vallette [2012]). The map of dg Hopf cooperads $\mathrm{Graphs}_n \to \mathrm{e}_n^c$ above is a quasi-isomorphism.

The advantage of the dg Hopf cooperad Graphs_n over the smaller weakly equivalent cooperad e_n^c is that it has a large group of automorphism. More precisely we may consider the dg Lie algebra

$$\mathbb{Q}L \ltimes \mathrm{GC}_n, \tag{10}$$

where the generator L on the left acts on a graph by multiplication with the loop order.

$$[L, \gamma] = (\#\mathrm{loops}) \cdot \gamma$$

The statement is now that the dg Lie algebra (10) acts on the Hopf cooperad Graphs_n by biderivations, i.e., compatibly with both the cooperad and the dg commutative algebra structures. More concretely, the action of the generator L on a graph $\Gamma \in \mathrm{Graphs}_n$ is by

$$L \cdot \Gamma = (\#(\mathrm{edges}) - \#(\mathrm{internal\ vertices}))\Gamma.$$

The action of $\gamma \in \mathrm{GC}_n$ on $\Gamma \in \mathrm{Graphs}_n$ can be defined combinatorially as the contraction of subgraphs of shape γ in Γ, see Willwacher [2015b] for more details and the explicit formula.

3.5 A sketch of a proof of Theorem 6. As an illustration of the connection between the graph complexes GC_n of Section 2 we sketch here a proof of our intrinsic formality theorem (Theorem 6 above). We deviate slightly from the original reference Fresse and Willwacher [2015], where a different approach was used, using Bousfield's obstruction theory. The proof proceeds along the following sequence of steps.

1. As for most algebraic objects, we can define a deformation complex $\mathrm{Def}(\mathfrak{C}, \mathfrak{D})$ for Hopf cooperads $\mathfrak{C}$ and $\mathfrak{D}$, governing maps between $\mathfrak{C}$ and $\mathfrak{D}$. More concretely, $\mathrm{Def}(\mathfrak{C}, \mathfrak{D})$ is a dg Lie or L_∞-algebra, whose Maurer-Cartan elements correspond to maps from a cofibrant replacement of $\mathfrak{C}$ to a fibrant replacement of $\mathfrak{D}$. In particular $H^2(\mathrm{Def}(\mathfrak{C}, \mathfrak{D}))$ is a space of potential obstructions to constructing such maps. Furthermore, given a Hopf operad map $f : \mathfrak{C} \to \mathfrak{D}$ and the corresponding Maurer-Cartan element α_f, the twisted dg Lie algebra $\mathrm{Def}(\mathfrak{C} \xrightarrow{f} \mathfrak{D}) := \mathrm{Def}(\mathfrak{C}, \mathfrak{D})^{\alpha_f}$ governs deformations of the map f. We refer to Fresse, Turchin, and Willwacher [2017b, sections 3, 5] for details.

   We are interested in particular in $\mathrm{Def}(\mathrm{e}_n^c) := \mathrm{Def}(\mathrm{e}_n^c \xrightarrow{\mathrm{id}} \mathrm{e}_n^c)$, governing automorphisms of $\mathrm{e}_n^c = H(\mathrm{D}_n)$. If $\mathfrak{D}$ is a Hopf cooperad with $H(\mathfrak{D}) = \mathrm{e}_n^c$, then $H^2(\mathrm{Def}(\mathrm{e}_n^c))$ appears as a space of potential obstructions of lifting the cohomology map $\mathrm{e}_n^c \xrightarrow{=} H(\mathfrak{C})$ to a weak equivalence of Hopf cooperads between e_n^c and $\mathfrak{C}$. Our goal henceforth is to understand the space $H^2(\mathrm{Def}(\mathrm{e}_n^c))$, obstructing the intrinsic formality of the little n-disks operad.

2. If $\mathfrak{g}$ is a dg Lie algebra acting on e_n^c, or a quasi-isomorphic object, we obtain a map $H(\mathfrak{g}) \to H(\mathrm{Def}(\mathrm{e}_n^c))[1]$. It turns out that in the case of $\mathfrak{g} = \mathbb{Q}L \ltimes \mathrm{GC}_n$ acting on $\mathrm{Graphs}_n \simeq \mathrm{e}_n^c$ as described above the resulting map

$$\mathbb{Q}L \oplus H(\mathrm{GC}_n) \to H(\mathrm{Def}(\mathrm{e}_n^c))[1]$$

   is an isomorphism. This means in particular that the space of (potential) obstructions to intrinsic formality is given precisely by $H^1(\mathrm{GC}_n)$.

3. By the degree counting result (6) we hence see that for $n \geq 3$ the only possible obstructions are given by the graph cohomology classes represented by multiples of the loop graphs (4) appearing in (5). These graphs live in degree 1 only if n is divisible by 4. The intrinsic formality statment hence follows for $n \geq 3$ not divisible by 4.

4. Suppose next that $n \geq 3$ is divisible by 4. One can check that the $O(n)$ action on E_n is such that conjugation with the involution $S \in O(n)$ flipping the sign of one of the coordinates amounts to a multiplication of graphs in GC_n by $(-1)^{\#\mathrm{loops}}$. One can hence conclude that under the additional requirement of the presence of an involution on the operad as in Theorem 6, the relevant obstructions to intrinsic formality lie in the even loop order part $H(\mathrm{GC}_n)^{\mathbb{Z}_2}$. Hence the one-loop graphs cannot contribute.

By a similar analysis one can also make statements about the homotopy automorphisms of e_n^c for $n \geq 3$. Infinitesimally, they are governed by $\mathbb{Q}L \ltimes H^0(\mathrm{GC}_n)$. The piece $\mathbb{Q}L$ corresponds to the "trivial" automorphisms $S_\lambda : \mathsf{e}_n^c \to \mathsf{e}_n^c$ which just rescale the Lie cobracket by the factor $\lambda \in \mathbb{Q}^\times$. By invoking again the degree counting result 6 we see that (for $n \geq 3$) $H^0(\mathrm{GC}_n) = 0$ if $n \neq 3 \mod 4$ and otherwise $H^0(\mathrm{GC}_n)$ is one-dimensional, spanned by L_n. Hence one can say that for $n \geq 3$ the rationalization of the little n-disks operad has non-trivial homotopy automorphisms only if $n = 4k + 3$, and then only a 1-parameter family of such. Note that this is in striking contrast to the case of $n = 2$, where the homotopy automorphisms form the infinite dimensional Grothendieck-Teichmüller group, cf. Fresse [n.d.] or Theorem 1.

4 Mapping spaces and long knots

Let us turn again to the computation of the rational homotopy type of the space of long knots. Using the result (3) from the Goodwillie-Weiss embedding calculus we see that the quantity to evaluate is the mapping space $\mathrm{Map}^h_{op}(E_m, E_n)$. We are interested in the rationalization of this space. By Fresse, Turchin, and Willwacher [2017b, Theorem 15 and Proposition 6.1] the rationalization is weakly equivalent to the space

$$\mathrm{Map}^h_{op}(E_m, E_n^{\mathbb{Q}}) \simeq \mathrm{Map}^h_{\text{Hopf-Op}^c}(\Omega_\sharp E_n, \Omega_\sharp E_m) \simeq \mathrm{Map}^h_{\text{Hopf-Op}^c}(\mathsf{e}_n^c, \mathsf{e}_m^c)$$

if $n - m \geq 3$. For the last equivalence one uses the formality result of the previous section.

The space on the right-hand side of the above equation can be expressed through purely combinatorial data, and shown to be weakly equivalent to the nerve (Maurer-Cartan space) of the Lie algebra of hairy graphs $\mathrm{HGC}_{m,n}$.

Theorem 11 (Fresse, Turchin, and Willwacher [ibid.]). For $n \geq m \geq 2$ we have that

$$\mathrm{Map}^h_{\text{Hopf-Op}^c}(\mathsf{e}_n^c, \mathsf{e}_m^c) \simeq MC_\bullet(\mathrm{HGC}_{m,n}) := MC(\mathrm{HGC}\hat{\otimes}\Omega_{poly}(\Delta^\bullet))$$

where the hairy graph complex $\mathrm{HGC}_{m,n}$ is equipped with the Lie algebra structure of Section 2.1.

In fact, the same statement continues to hold for $m = 1$, if one modifies the Lie algebra structure on $\mathrm{HGC}_{1,n}$ to an L_∞-algebra structure described in Willwacher [2015a].

Theorem 11 states in particular that the real homotopy type of the spaces of long knots in codimension $n - m \geq 3$ is fully expressed through combinatorial data encoded in the graph complexes. In particular one finds that for $k \geq 0$ and $n-m \geq 3$ (see also Arone and Turchin [2015] for an earlier but weaker result)

$$\mathbb{Q} \otimes \pi_k(\overline{\mathrm{Emb}}_\partial(\mathbb{R}^m, \mathbb{R}^n)) \cong H^{-k-m}(\mathrm{HGC}_{m,n}).$$

Since the cohomology of the graph complex $\mathrm{HGC}_{m,n}$ can be explicitly computed in low degrees as described in Section 2.1, this result allows us to compute rational homotopy groups for low degrees k explicitly. As one application, let us consider here the case $k = 0$, detecting nontrivial knottings of $\mathbb{R}^m$ in $\mathbb{R}^n$. From the degree bound (8) we conclude that the only possible contribution can come from graphs in loop order 0, and those graph cohomology classes are listed in (7). Concretely the "line" graph of degree $n-m-2$ occurring for $n-m$ even produces nontrivial knottings if $n = 4j-1$ and $m = 2j$ for some j. Similarly the tripod graph of degree $2(n-m)-3$ occurring for $n-m$ odd yields nontrivial zeroth homotopy $n = 6j$ and $m = 4j-1$ for some j, see Fresse, Turchin, and Willwacher [2017b, Corollary 20]. In particular one hence recovers Haefliger's classical result Haefliger [1965] on the existence of non-trivial knots of dimension $4j-1$ in $6j$-space.

5 Outlook, extensions and open problems

Above we have seen that a rational version of the Goodwillie-Weiss manifold calculus can be used to compute the rational homotopy type of the space of higher dimensional long knots through the combinatorial structure of the graph complexes $\mathrm{HGC}_{m,n}$. Evidently, one can similarly hope to attack arbitrary embedding spaces $\mathrm{Emb}(M,N)$ via the rationalized Goodwillie-Weiss calculus. To complete this program one in particular needs to understand the rational homotopy types of all spaces and operads involved. As of today, at least over the reals, we understand the real homotopy types of (unframed) configuration spaces of points on compact orientable manifolds through the works Campos and Willwacher [2016] and Idrissi [2016]. Also, we understand the real homotopy type of the framed little disks operads Khoroshkin and Willwacher [2017]. The main open problem is the following.

Open Problem 12. Determine the real or rational homotopy type of the n-framed configuration spaces of points on a manifold M as a right E_n^{fr}-module, where $n \leq \dim M$.

This problem is also closely related to Open Open Problem 9 above.

After the real or rational homotopy types of the aforementioned objects are understood there should be no principle obstacle for generalizing the mapping space

space computations we reviewed here to arbitrary M and N. I can even describe the expected form of the graph complexes replacing the hairy graph complexes $\mathrm{HGC}_{m,n}$ in this setting. The relevant graphs should be hairy graphs as before, except that the hairs are additionaly decorated by a dgca model for the source space M, and the internal vertices may be decorated by the homology of the target N as in Campos and Willwacher [2016].

Unfortunately, the embedding calculus is fully applicable only in codimensions $\dim N - \dim M \geq 3$. In lower codimension one still has the map (2) (from left to right), but one can not assert it to be a weak equivalence. Nevertheless I expect that valuable information about the left-hand side can be obtained from the right-hand side. For example, in codimension 0, we have a map into the homotopy automorphisms of right framed-little disks (co)modules

$$\mathrm{Diff}(M) \to \mathrm{Aut}^h_{\Omega(E_n^{fr})\text{-comod}}(\Omega(\mathrm{conf}_M)). \tag{11}$$

I claim that the right-hand side can be computed and expressed through graph complexes. This then provides an arena in which one can study diffeomorphism groups, although the map (11) is generally not (expected to be) a rational homotopy equivalence.

References

Vladimir Arnold (1969). "The cohomology ring of the group of dyed braids". Mat. Zametki 5, pp. 227–231 (cit. on p. 1269).

G. Arone and V. Turchin (2014). "On the rational homology of high-dimensional analogues ofspaces of long knots". Geom. Topol. 18.3, pp. 1261–1322 (cit. on p. 1268).

– (2015). "Graph-complexes computing the rational homotopy of high dimensionalanalogues of spaces of long knots". Ann. Inst. Fourier 65.1, pp. 1–62 (cit. on pp. 1269, 1278).

Francis C. S. Brown (2012). "Mixed Tate Motives over Spec($\mathbb{Z}$)". Annals of Math. 175.2, pp. 949–976 (cit. on p. 1265).

Ricardo Campos and Thomas Willwacher (2016). "A model for configuration spaces of points". arXiv: 1604.02043 (cit. on pp. 1261, 1278, 1279).

Fred R. Cohen (1976). "The homology of C_{n+1}-spaces, $n \geq 0$". Lecture Notes in Mathematics 533, pp. 207–351 (cit. on p. 1269).

J. Conant, J. Costello, V. Turchin, and P. Weed (2014). "Two-loop part of the rational homotopy of spaces of long embeddings". J. Knot Theory Ramifications 23.4, p. 1450018 (cit. on p. 1268).

V. G. Drinfeld (1990). “On quasitriangular quasi-Hopf algebras and on a group that isclosely connected with Gal($\overline{\mathrm{Q}}$/Q)”. Russian. Algebra i Analiz 2.4, pp. 149–181 (cit. on p. 1265).

J. Ducoulombier and V. Turchin (2017). “Delooping manifold calculus tower on a closed disc”. arXiv: 1708.02203 (cit. on p. 1261).

W. Dwyer and K. Hess (2012). “Long knots and maps between operads”. Geom. Topol. 16.2, pp. 919–955 (cit. on p. 1261).

B. Fresse, V. Turchin, and T. Willwacher (2017a). “The homotopy theory of operad subcategories”. To appear in the Journal of Homotopy and Related Structures. arXiv: 1703.06093 (cit. on p. 1271).

B. Fresse and T. Willwacher (2015). “The intrinsic formality of E_n-operads”. arXiv: 1503.08699 (cit. on pp. 1261, 1270, 1272, 1275).

Benoit Fresse (n.d.). Homotopy of Operads and Grothendieck-Teichmüller Groups. Mathematical Surveys and Monographs. American Mathematical Society (cit. on pp. 1262, 1270, 1271, 1277).

Benoit Fresse, Victor Turchin, and Thomas Willwacher (2017b). “The rational homotopy of mapping spaces of E_n operads”. arXiv: 1703.06123 (cit. on pp. 1261, 1269, 1276–1278).

Ezra Getzler and J. D. S. Jones (1994). “Operads, homotopy algebra and iterated integrals for double loop spaces”. arXiv: hep-th/9403055 (cit. on p. 1271).

Jeffrey Giansiracusa and Paolo Salvatore (2010). “Formality of the framed little 2-discs operad and semidirectproducts”. Contemp. Math. 519, pp. 115–121. MR: 2648708 (cit. on p. 1272).

F. Guillén Santos, V. Navarro, P. Pascual, and A. Roig (2005). “Moduli spaces and formal operads”. Duke Math. J. 129.2, pp. 291–335. MR: 2165544 (cit. on p. 1271).

André Haefliger (1965). “Knotted $(4k-1)$-spheres in $6k$-space”. Ann. Math. 75.2, pp. 452–466 (cit. on p. 1278).

Najib Idrissi (2016). “The Lambrechts–Stanley Model of Configuration Spaces”. arXiv: 1608.08054 (cit. on pp. 1261, 1278).

A. Khoroshkin, T. Willwacher, and M. Živković (2015). “Differentials on graph complexes II - Hairy Graphs”. arxiv:1508.01281 (cit. on p. 1269).

– (2017). “Differentials on graph complexes”. Adv. Math. 307, pp. 1184–1214 (cit. on p. 1265).

Anton Khoroshkin and Thomas Willwacher (2017). “Real models for the framed little n-disks operads ”. arXiv: 1705.08108 (cit. on pp. 1273, 1278).

Jan A. Kneissler (2000). “On spaces of connected graphs. I. Properties of ladders”. Ser. Knots Everything 24, pp. 252–273 (cit. on p. 1265).

– (2001a). "On spaces of connected graphs. II. Relations in thealgebra Λ". J. Knot Theory Ramifications 10.5. Knots in Hellas '98, Vol. 3 (Delphi), pp. 667–674 (cit. on p. 1265).

– (2001b). "On spaces of connected graphs. III. The ladder filtration". J. Knot Theory Ramifications 10.5. Knots in Hellas '98, Vol. 3 (Delphi), pp. 675–686 (cit. on p. 1265).

Maxim Kontsevich (1993). "Formal (non)commutative symplectic geometry", pp. 173–187 (cit. on p. 1264).

– (1999). "Operads and motives in deformation quantization". Lett. Math. Phys. 48.1. Moshé Flato (1937–1998), pp. 35–72 (cit. on pp. 1271, 1273, 1275).

J.-L. Loday and B. Vallette (2012). Algebraic Operads. Grundlehren der mathematischen Wissenschaften 346. Berlin: Springer (cit. on pp. 1262, 1272, 1275).

J. P. May (1972). The geometry of iterated loop spaces. Lectures Notes in Mathematics, Vol. 271. Springer-Verlag, Berlin-New York, pp. viii+175 (cit. on p. 1262).

Syunji Moriya (2017). "Non-Formality of the Odd Dimensional Framed Little Balls Operads". International Mathematics Research Notices (cit. on p. 1273).

"P. Etingof's conjecture about Drinfeld associators" (2014). arXiv: 1404.2047 (cit. on p. 1265).

Pavol Ševera (2010). "Formality of the chain operad of framed little disks". Lett. Math. Phys. 93.1, pp. 29–35. MR: 2661521 (cit. on p. 1272).

Dmitry E. Tamarkin (2003). "Formality of chain operad of little discs". Lett. Math. Phys. 66.1-2, pp. 65–72. MR: 2064592 (cit. on p. 1271).

Victor Turchin and Thomas Willwacher (2014). "Relative (non-)formality of the little cubes operads and the algebraic Cerf Lemma". arXiv: 1409.0163 (cit. on p. 1272).

Pierre Vogel (2011). "Algebraic structures on modules of diagrams". J. Pure Appl. Algebra 215.6, pp. 1292–1339 (cit. on p. 1265).

Michael Weiss (1999). "Embeddings from the point of view of immersion theory. I". Geom. Topol. 3, 67–101 (electronic) (cit. on p. 1260).

Michael Weiss and Pedro Boavida de Brito (2013). "Manifold calculus and homotopy sheaves". Homology Homotopy Appl. 15.2, pp. 361–383 (cit. on p. 1260).

– (2015). "Spaces of smooth embeddings and configuration categories". arXiv: 1502.01640 (cit. on p. 1261).

Michael Weiss and Thomas G. Goodwillie (1999). "Embeddings from the point of view of immersion theory. II". Geom. Topol. 3, 103–118 (electronic) (cit. on p. 1260).

Thomas Willwacher (2015a). "Deformation quantization and the Gerstenhaber structure on the homology of knot spaces". arXiv: 1506.07078 (cit. on p. 1277).

Thomas Willwacher (2015b). "M. Kontsevich's graph complex and the Grothendieck-Teichmüller Lie algebra". Invent. Math. 200.3, pp. 671–760 (cit. on pp. 1264, 1265, 1268, 1275).

Thomas Willwacher and Marko Živković (2015). "Multiple edges in M. Kontsevich's graph complexes andcomputations of the dimensions and Euler characteristics". Adv. Math. 272, pp. 553–578. MR: 3303240 (cit. on p. 1267).

Received 2018-02-27.

Thomas Willwacher
Department of Mathematics
ETH Zurich
Rämistrasse 101
8092 Zurich
Switzerland
thomas.willwacher@math.ethz.ch

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (1281–1300)

REPRESENTATION THEORY OF W-ALGEBRAS AND HIGGS BRANCH CONJECTURE

TOMOYUKI ARAKAWA (荒川知幸)

Abstract

We survey a number of results regarding the representation theory of W-algebras and their connection with the resent development of the four dimensional $N = 2$ superconformal field theories.

1 Introduction

(Affine) W-algebras appeared in 80's in the study of the two-dimensional conformal field theory in physics. They can be regarded as a generalization of infinite-dimensional Lie algebras such as affine Kac-Moody algebras and the Virasoro algebra, although W-algebras are not Lie algebras but vertex algebras in general. W-algebras may be also considered as an affinizaton of finite W-algebras introduced by Premet [2002] as a natural quantization of Slodowy slices. W-algebras play important roles not only in conformal field theories but also in integrable systems (e.g. V. G. Drinfeld and Sokolov [1984], De Sole, V. G. Kac, and Valeri [2013], and Bakalov and Milanov [2013]), the geometric Langlands program (e.g. E. Frenkel [2007], Gaitsgory [2016], Tan [2016], and Aganagic, E. Frenkel, and Okounkov [2017]) and four-dimensional gauge theories (e.g. Alday, Gaiotto, and Tachikawa [2010], Schiffmann and Vasserot [2013], Maulik and Okounkov [2012], and Braverman, Finkelberg, and Nakajima [2016a]).

In this note we survey the resent development of the representation theory of W-algebras. One of the fundamental problems in W-algebras was the Frenkel-Kac-Wakimoto conjecture E. Frenkel, V. Kac, and Wakimoto [1992] that stated the existence and construction of rational W-algebras, which generalizes the integrable representations of affine Kac-Moody algebras and the minimal series representations of the Virasoro algebra. The notion of the associated varieties of vertex algebras played a crucial role in the proof of the Frenkel-Kac-Wakimoto conjecture, and has revealed an unexpected connection of vertex algebras with the geometric invariants called the Higgs branches in the four dimensional $N = 2$ superconformal field theories.

MSC2010: primary 81R10; secondary 17B67, 17B67.

Acknowledgments. The author benefited greatly from discussion with Christopher Beem, Davide Gaiotto, Madalena Lemos, Victor Kac, Anne Moreau, Hiraku Nakajima, Takahiro Nishinaka, Wolfger Peelaers, Leonardo Rastelli, Shu-Heng Shao, Yuji Tachikawa, and Dan Xie. The author is partially supported in part by JSPS KAKENHI Grant Numbers 17H01086, 17K18724.

2 Vertex algebras

A *vertex algebra* consists of a vector space V with a distinguished vacuum vector $|0\rangle \in V$ and a vertex operation, which is a linear map $V \otimes V \to V((z))$, written $a \otimes b \mapsto a(z)b = (\sum_{n\in\mathbb{Z}} a_{(n)} z^{-n-1})b$, such that the following are satisfied:

- (Unit axioms) $(|0\rangle)(z) = 1_V$ and $a(z)|0\rangle \in a + zV[[z]]$ for all $a \in V$.
- (Locality) $(z-w)^n [a(z), b(w)] = 0$ for a sufficiently large n for all $a, b \in V$.

The operator $T : a \mapsto a_{(-2)}|0\rangle$ is called the translation operator and it satisfies $(Ta)(z) = [T, a(z)] = \partial_z a(z)$. The operators $a_{(n)}$ are called *modes*.

For elements a, b of a vertex algebra V we have the following *Borcherds identity* for any $m, n \in \mathbb{Z}$:

$$(1)\qquad [a_{(m)}, b_{(n)}] = \sum_{j\geq 0} \binom{m}{j} (a_{(j)}b)_{(m+n-j)},$$

$$(2)\qquad (a_{(m)}b)_{(n)} = \sum_{j\geq 0} (-1)^j \binom{m}{j} (a_{(m-j)}b_{(n+j)} - (-1)^m b_{(m+n-j)}a_{(j)}).$$

By regarding the Borcherds identity as fundamental relations, representations of a vertex algebra are naturally defined (see V. Kac [1998] and E. Frenkel and Ben-Zvi [2004] for the details).

One of the basic examples of vertex algebras are universal affine vertex algebras. Let G be a simply connected simple algebraic group, $\mathfrak{g} = \mathrm{Lie}(G)$. Let $\widehat{\mathfrak{g}} = \mathfrak{g}[t, t^{-1}] \oplus \mathbb{C}K$ be the affine Kac-Moody algebra associated with $\mathfrak{g}$. The commutation relations of $\widehat{\mathfrak{g}}$ are given by

$$(3)\quad [xt^m, yt^n] = [x, y]t^{m+n} + m\delta_{m+n,0}(x|y)K, \quad [K, \widehat{\mathfrak{g}}] = 0 \quad (x, y \in \mathfrak{g},\ m, n \in \mathbb{Z}),$$

where $(\ |\ )$ is the normalized invariant inner product of $\mathfrak{g}$, that is, $(\ |\ ) = 1/2h^\vee \times$Killing form and $h^\vee$ is the dual Coxeter number of $\mathfrak{g}$. For $k \in \mathbb{C}$, let

$$V^k(\mathfrak{g}) = U(\widehat{\mathfrak{g}}) \otimes_{U(\mathfrak{g}[t]\oplus\mathbb{C}K)} \mathbb{C}_k,$$

where $\mathbb{C}_k$ is the one-dimensional representation of $\mathfrak{g}[t] \oplus \mathbb{C}K$ on which $\mathfrak{g}[t]$ acts trivially and K acts as multiplication by k. There is a unique vertex algebra structure on $V^k(\mathfrak{g})$ such that $|0\rangle = 1 \otimes 1$ is the vacuum vector and

$$x(z) = \sum_{n \in \mathbb{Z}} (xt^n) z^{-n-1} \quad (x \in \mathfrak{g}).$$

Here on the left-hand-side $\mathfrak{g}$ is considered as a subspace of $V^k(\mathfrak{g})$ by the embedding $\mathfrak{g} \hookrightarrow V^k(\mathfrak{g})$, $x \mapsto (xt^{-1})|0\rangle$. $V^k(\mathfrak{g})$ is called the *universal affine vertex algebra associated with* $\mathfrak{g}$ *at a level* k. The Borcherds identity (1) for $x, y \in \mathfrak{g} \subset V^k(\mathfrak{g})$ is identical to the commutation relation (3) with $K = k$ id, and hence, any $V^k(\mathfrak{g})$-module is a $\widehat{\mathfrak{g}}$-module of level k. Conversely, any smooth $\widehat{\mathfrak{g}}$-module of level k is naturally a $V^k(\mathfrak{g})$-module, and therefore, the category $V^k(\mathfrak{g})$ -Mod of $V^k(\mathfrak{g})$-modules is the same as the category of smooth $\widehat{\mathfrak{g}}$-modules of level k. Let $L_k(\mathfrak{g})$ be the unique simple graded quotient of $V^k(\mathfrak{g})$, which is isomorphic to the irreducible highest weight representation $L(k\Lambda_0)$ with highest weight $k\Lambda_0$ as a $\widehat{\mathfrak{g}}$-module. The vertex algebra $L_k(\mathfrak{g})$ is called the *simple affine vertex algebra* associated with $\mathfrak{g}$ at level k, and $L_k(\mathfrak{g})$ -Mod forms a full subcategory of $V^k(\mathfrak{g})$ -Mod, the category of smooth $\widehat{\mathfrak{g}}$-modules of level k.

A vertex algebra V is called *commutative* if both sides of (1) are zero for all $a, b \in V$, $m, n \in \mathbb{Z}$. If this is the case, V can be regarded as a *differential algebra* (=a unital commutative algebra with a derivation) by the multiplication $a.b = a_{(-1)}b$ and the derivation T. Conversely, any differential algebra can be naturally equipped with the structure of a commutative vertex algebra. Hence, commutative vertex algebras are the same[1] as differential algebras (Borcherds [1986]).

Let X be an affine scheme, $J_\infty X$ the *arc space* of X that is defined by the functor of points $\mathrm{Hom}(\mathrm{Spec}\, R, J_\infty X) = \mathrm{Hom}(\mathrm{Spec}\, R[[t]], X)$. The ring $\mathbb{C}[J_\infty X]$ is naturally a differential algebra, and hence is a commutative vertex algebra. In the case that X is a Poisson scheme $\mathbb{C}[J_\infty X]$ has the structure of *Poisson vertex algebra* (Arakawa [2012b]), which is a vertex algebra analogue of Poisson algebra (see E. Frenkel and Ben-Zvi [2004] and V. Kac [2015] for the precise definition).

It is known by Haisheng Li [2005] that any vertex algebra V is canonically filtered, and hence can be regarded[2] as a quantization of the associated graded Poisson vertex algebra $\mathrm{gr}\, V = \bigoplus_p F^pV/F^{p+1}V$, where $F^\bullet V$ is the canonical filtration of V. By definition,

$$F^pV = \mathrm{span}_{\mathbb{C}}\{(a_1)_{(-n_1-1)} \dots (a_r)_{(-n_r-1)}|0\rangle \mid a_i \in V,\ n_i \geq 0,\ \sum_i n_i \geq p\}.$$

[1]However, the modules of a commutative vertex algebra are not the same as the modules as a differential algebra.

[2]This filtration is separated if V is non-negatively graded, which we assume.

The subspace

$$R_V := V/F^1V = F^0V/F^1V \subset \operatorname{gr} V$$

is called *Zhu's C_2-algebra of V*. The Poisson vertex algebra structure of gr V restricts to the Poisson algebra structure of R_V, which is given by

$$\bar{a}.\bar{b} = \overline{a_{(-1)}b}, \quad \{\bar{a}, \bar{b}\} = \overline{a_{(0)}b}.$$

The Poisson variety

$$X_V = \operatorname{Specm}(R_V)$$

called the *associated variety* of V (Arakawa [2012b]). We have Li [2005] the inclusion

(4) $$\operatorname{Specm}(\operatorname{gr} V) \subset J_\infty X_V.$$

A vertex algebra V is called *finitely strongly generated* if R_V is finitely generated. In this note all vertex algebras are assumed to be finitely strongly generated. V is called *lisse* (or *C_2-cofinite*) if $\dim X_V = 0$. By (4), it follows that V is lisse if and only if $\dim \operatorname{Spec}(\operatorname{gr} V) = 0$ (Arakawa [2012b]). Hence lisse vertex algebras can be regarded as an analogue of finite-dimensional algebras.

For instance, consider the case $V = V^k(\mathfrak{g})$. We have $F^1V^k(\mathfrak{g})) = \mathfrak{g}[t^{-1}]t^{-2}V^k(\mathfrak{g})$, and there is an isomorphism of Poisson algebras

$$\mathbb{C}[\mathfrak{g}^*] \xrightarrow{\sim} R_V, \quad x_1 \dots x_r \mapsto \overline{(x_1t^{-1}) \dots (x_rt^{-1})|0\rangle} \quad (x_i \in \mathfrak{g}).$$

Hence

(5) $$X_{V^k(\mathfrak{g})} \cong \mathfrak{g}^*.$$

Also, we have the isomorphism $\operatorname{Spec}(\operatorname{gr} V^k(\mathfrak{g})) \cong J_\infty \mathfrak{g}^*$. By (5), we have $X_{L_k(\mathfrak{g})} \subset \mathfrak{g}^*$, which is G-invariant and conic. It is known Dong and Mason [2006] that

(6) $$L_k(\mathfrak{g}) \text{ is lisse} \iff L_k(\mathfrak{g}) \text{ is integrable as a } \widehat{\mathfrak{g}}\text{-module } (\iff k \in \mathbb{Z}_{\geq 0}).$$

Hence the lisse condition may be regarded as a generalization of the integrability condition to an arbitrary vertex algebra.

A vertex algebra is called *conformal* if there exists a vector ω, called the *conformal vector*, such that the corresponding field $\omega(z) = \sum_{n \in \mathbb{Z}} L_n z^{-n-2}$ satisfies the following conditions. (1) $[L_m, L_n] = (m-n)L_{m+n} + \frac{m^3-m}{12}\delta_{m+n,0} c \operatorname{id}_V$, where c is a constant called the central charge of V; (2) L_0 acts semisimply on V; (3) $L_{-1} = T$. For a conformal vertex algebra V we set $V_\Delta = \{v \in V \mid L_0 v = \Delta V\}$, so that $V = \bigoplus_\Delta V_\Delta$. The universal affine vertex algebra $V^k(\mathfrak{g})$ is conformal by the Sugawara construction provided that $k \neq -h^\vee$. A *positive energy representation* M of a conformal vertex algebra V is a V-module M on which L_0 acts diagonally and the L_0-eivenvalues on M is bounded from below. An *ordinary representation* is a positive energy representation such that each L_0-eivenspaces are finite-dimensional. For a finitely generated ordinary representation M, the normalized character

$$\chi_V(q) = \mathrm{tr}_V(q^{L_0 - c/24})$$

is well-defined.

For a conformal vertex algebra $V = \bigoplus_\Delta V_\Delta$, one defines *Zhu's algebra* I. B. Frenkel and Zhu [1992] $\mathrm{Zhu}(V)$ of V by

$$\mathrm{Zhu}(V) = V/V \circ V, \quad V \circ V = \mathrm{span}_{\mathbb{C}}\{a \circ b \mid a, b \in V\},$$

where $a \circ b = \sum_{i \geq 0} \binom{\Delta_a}{i} a_{(i-2)} b$ for $a \in \Delta_a$. $\mathrm{Zhu}(V)$ is a unital associative algebra by the multiplication $a * b = \sum_{i \geq 0} \binom{\Delta_a}{i} a_{(i-1)} b$. There is a bijection between the isomorphism classes $\mathrm{Irrep}(V)$ of simple positive energy representation of V and that of simple $\mathrm{Zhu}(V)$-modules (I. B. Frenkel and Zhu [1992] and Zhu [1996]). The grading of V gives a filtration on $\mathrm{Zhu}(V)$ which makes it quasi-commutative, and there is a surjective map

$$R_V \twoheadrightarrow \mathrm{gr}\,\mathrm{Zhu}(V) \tag{7}$$

of Poisson algebras. Hence, if V is lisse then $\mathrm{Zhu}(V)$ is finite-dimensional, so there are only finitely many irreducible positive energy representations of V. Moreover, the lisse condition implies that any simple V-module is a positive energy representation (Abe, Buhl, and Dong [2004]).

A conformal vertex algebra is called *rational* if any positive energy representation of V is completely reducible. For instance, the simple affine vertex algebra $L_k(\mathfrak{g})$ is rational if and only if $L_k(\mathfrak{g})$ is integrable, and if this is the case $L_k(\mathfrak{g})$ -Mod is exactly the category of integrable representations of $\widehat{\mathfrak{g}}$ at level k. A theorem of Y. Zhu [1996] states that if V is a rational, lisse, $\mathbb{Z}_{\geq 0}$-graded conformal vertex algebra such that $V_0 = \mathbb{C}|0\rangle$, then the character $\chi_M(e^{2\pi i \tau})$ converges to a holomorphic functor on the upper half plane for any $M \in \mathrm{Irrep}(V)$. Moreover, the space spanned by the characters $\chi_M(e^{2\pi i \tau})$, $M \in \mathrm{Irrep}(V)$, is invariant under the natural action of $SL_2(\mathbb{Z})$. This theorem was strengthened in Dong, Lin, and Ng [2015] to the fact that $\{\chi_M(e^{2\pi i \tau}) \mid M \in \mathrm{Irrep}(V)\}$ forms a vector valued modular function by showing the congruence property. Furthermore, it has been shown in Huang [2008] that the category of V-modules form a modular tensor category.

3 W-algebras

W-algebras are defined by the method of the *quantized Drinfeld-Sokolov reduction* that was discovered by Feigin and E. Frenkel [1990]. In the most general definition of W-algebras given by V. Kac, Roan, and Wakimoto [2003], W-algebras are associated with the pair $(\mathfrak{g}, f)$ of a simple Lie algebra $\mathfrak{g}$ and a nilpotent element $f \in \mathfrak{g}$. The corresponding W-algebra is a one-parameter family of vertex algebra denoted by $\mathcal{W}^k(\mathfrak{g}, f)$, $k \in \mathbb{C}$. By definition,

$$\mathcal{W}^k(\mathfrak{g}, f) := H^0_{DS,f}(V^k(\mathfrak{g})),$$

where $H^{\bullet}_{DS,f}(M)$ denotes the BRST cohomology of the quantized Drinfeld-Sokolov reduction associated with $(\mathfrak{g}, f)$ with coefficient in a $V^k(\mathfrak{g})$-module M, which is defined as follows. Let $\{e, h, f\}$ be an $\mathfrak{sl}_2$-triple associated with f, $\mathfrak{g}_j = \{x \in \mathfrak{g} \mid [h, x] = 2j\}$, so that $\mathfrak{g} = \bigoplus_{j \in \frac{1}{2}\mathbb{Z}} \mathfrak{g}_j$. Set $\mathfrak{g}_{\geq 1} = \bigoplus_{j \geq 1} \mathfrak{g}_j$, $\mathfrak{g}_{>0} = \bigoplus_{j \geq 1/2} \mathfrak{g}_j$. Then $\chi : \mathfrak{g}_{\geq 1}[t, t^{-1}] \to \mathbb{C}$, $xt^n \mapsto \delta_{n,-1}(f|x)$, defines a character. Let $F_\chi = U(\mathfrak{g}_{>0}[t, t^{-1}]) \otimes_{U(\mathfrak{g}_{>0}[t] + \mathfrak{g}_{\geq 1}[t, t^{-1}])} \mathbb{C}_\chi$, where $\mathbb{C}_\chi$ is the one-dimensional representation of $\mathfrak{g}_{>0}[t] + \mathfrak{g}_{\geq 1}[t, t^{-1}]$ on which $\mathfrak{g}_{\geq 1}[t, t^{-1}]$ acts by the character χ and $\mathfrak{g}_{>0}[t]$ acts triviality. Then, for a $V^k(\mathfrak{g})$-module M,

$$H^{\bullet}_{DS,f}(M) = H^{\frac{\infty}{2}+\bullet}(\mathfrak{g}_{>0}[t, t^{-1}], M \otimes F_\chi),$$

where $H^{\frac{\infty}{2}+\bullet}(\mathfrak{g}_{>0}[t, t^{-1}], N)$ is the semi-infinite $\mathfrak{g}_{>0}[t, t^{-1}]$-cohomology Feĭgin [1984] with coefficient in a $\mathfrak{g}_{>0}[t, t^{-1}]$-module N. Since it is defined by a BRST cohomology, $\mathcal{W}^k(\mathfrak{g}, f)$ is naturally a vertex algebra, which is called *W-algebra associated with* $(\mathfrak{g}, f)$ *at level* k. By E. Frenkel and Ben-Zvi [2004] and V. G. Kac and Wakimoto [2004], we know that $H^i_{DS,f}(V^k(\mathfrak{g})) = 0$ for $i \neq 0$. If $f = 0$ we have by definition $\mathcal{W}^k(\mathfrak{g}, f) = V^k(\mathfrak{g})$. The W-algebra $\mathcal{W}^k(\mathfrak{g}, f)$ is conformal provided that $k \neq -h^\vee$.

Let $\mathcal{S}_f = f + \mathfrak{g}^e \subset \mathfrak{g} \cong \mathfrak{g}^*$, the *Slodowy slice* at f, where $\mathfrak{g}^e$ denotes the centralizer of e in $\mathfrak{g}$. The affine variety $\mathcal{S}_f$ has a Poisson structure obtained from that of $\mathfrak{g}^*$ by Hamiltonian reduction (Gan and Ginzburg [2002]). We have

$$X_{\mathcal{W}^k(\mathfrak{g},f)} \cong \mathcal{S}_f, \quad \operatorname{Spec}(\operatorname{gr} \mathcal{W}^k(\mathfrak{g}, f)) \cong J_\infty \mathcal{S}_f \tag{8}$$

(De Sole and V. G. Kac [2006] and Arakawa [2015a]). Also, we have

$$\operatorname{Zhu}(\mathcal{W}^k(\mathfrak{g}, f)) \cong U(\mathfrak{g}, f)$$

(Arakawa [2007] and De Sole and V. G. Kac [2006]), where $U(\mathfrak{g}, f)$ is the *finite W-algebra* associated with $(\mathfrak{g}, f)$ (Premet [2002]). Therefore, the W-algebra $\mathcal{W}^k(\mathfrak{g}, f)$ can be regarded as an affinization of the finite W-algebra $U(\mathfrak{g}, f)$. The map (7) for $\mathcal{W}^k(\mathfrak{g}, f)$

is an isomorphism, which recovers the fact Premet [2002] and Gan and Ginzburg [2002] that $U(\mathfrak{g}, f)$ is a quantization of the Slodowy slice $\mathcal{S}_f$. The definition of $\mathcal{W}^k(\mathfrak{g}, f)$ naturally extends V. Kac, Roan, and Wakimoto [2003] to the case that $\mathfrak{g}$ is a (basic classical) Lie superalgebra and f is a nilpotent element in the even part of $\mathfrak{g}$.

We have $\mathcal{W}^k(\mathfrak{g}, f) \cong \mathcal{W}^k(\mathfrak{g}, f')$ if f and f' belong to the same nilpotent orbit of $\mathfrak{g}$. The W-algebra associated with a minimal nilpotent element f_{mim} and a principal nilpotent element f_{prin} are called a minimal W-algebra and a principal W-algebra, respectively. For $\mathfrak{g} = \mathfrak{sl}_2$, these two coincide and are isomorphic to the Virasoro vertex algebra of central charge $1-6(k-1)^2/(k+2)$ provided that $k \neq -2$. In V. Kac, Roan, and Wakimoto [ibid.] it was shown that almost every superconformal algebra appears as the minimal W-algebra $\mathcal{W}^k(\mathfrak{g}, f_{min})$ for some Lie superalgebra $\mathfrak{g}$, by describing the generators and the relations (OPEs) of minimal W-algebras. Except for some special cases, the presentation of $\mathcal{W}^k(\mathfrak{g}, f)$ by generators and relations is not known for other nilpotent elements.

Historically, the principal W-algebras were first extensively studied (see Bouwknegt and Schoutens [1995]). In the case that $\mathfrak{g} = \mathfrak{sl}_n$, the non-critical principal W-algebras is isomorphic to the Fateev-Lukyanov's W_n-algebra Fateev and Lykyanov [1988] (Feigin and E. Frenkel [1990] and E. Frenkel and Ben-Zvi [2004]). The critical principal W-algebra $\mathcal{W}^{-h^\vee}(\mathfrak{g}, f_{prin})$ is isomorphic to the Feigin-Frenkel center $\mathfrak{z}(\widehat{\mathfrak{g}})$ of $\widehat{\mathfrak{g}}$, that is the center of the critical affine vertex algebra $V^{-h^\vee}(\mathfrak{g})$ (Feigin and E. Frenkel [1992]). For a general f, we have the isomorphism

$$\mathfrak{z}(\widehat{\mathfrak{g}}) \cong Z(\mathcal{W}^{-h^\vee}(\mathfrak{g}, f)),$$

(Arakawa [2012a] and Arakawa and Moreau [2018a]), where $Z(\mathcal{W}^{-h^\vee}(\mathfrak{g}, f))$ denotes the center of $\mathcal{W}^{-h^\vee}(\mathfrak{g}, f)$. This fact has an application to *Vinberg's Problem* for the centralizer $\mathfrak{g}^e$ of e in $\mathfrak{g}$ Arakawa and Premet [2017].

4 Representation theory of W-algebras

The definition of $\mathcal{W}^k(\mathfrak{g}, f)$ by the quantized Drinfeld-Sokolov reduction gives rise to a functor

$$V^k(\mathfrak{g})\text{-Mod} \rightarrow \mathcal{W}^k(\mathfrak{g}, f)\text{-Mod}$$
$$M \mapsto H^0_{DS,f}(M).$$

Let $\mathcal{O}_k$ be the category $\mathcal{O}$ of $\widehat{\mathfrak{g}}$ at level k. Then $\mathcal{O}_k$ is naturally considered as a full subcategory of $V^k(\mathfrak{g})$-Mod. For a weight λ of $\widehat{\mathfrak{g}}$ of level k, let $L(\lambda)$ be the irreducible highest weight representations of $\widehat{\mathfrak{g}}$ with highest weight λ.

Theorem 1 (Arakawa [2005]). *Let $f_{min} \in \mathbb{O}_{min}$ and let k be an arbitrary complex number.*

i. *We have $H^i_{DS,f_{min}}(M) = 0$ for any $M \in \mathcal{O}_k$ and $i \in \mathbb{Z}\backslash\{0\}$. Therefore, the functor*

$$\mathcal{O}_k \to \mathcal{W}^k(\mathfrak{g}, f_{min})\text{-Mod}, \quad M \mapsto H^0_{DS,f_{min}}(M)$$

is exact.

ii. *For a weight λ of $\widehat{\mathfrak{g}}$ of level k, $H^0_{DS,f_{min}}(L(\lambda))$ is zero or isomorphic to an irreducible highest weight representation of $\mathcal{W}^k(\mathfrak{g}, f_{min})$. Moreover, any irreducible highest weight representation the minimal W-algebra $\mathcal{W}^k(\mathfrak{g}, f_{min})$ arises in this way.*

By Theorem 1 and the Euler-Poincaré principal, the character $\operatorname{ch} H^0_{DS,f_\theta}(L(\lambda))$ is expressed in terms of $\operatorname{ch} L(\lambda)$. Since $\operatorname{ch} L(\lambda)$ is known Kashiwara and Tanisaki [2000] for all non-critical weight λ, Theorem 1 determines the character of all non-critical irreducible highest weight representation of $\mathcal{W}^k(\mathfrak{g}, f_{min})$. In the case that k is critical the character of irreducible highest weight representation of $\mathcal{W}^k(\mathfrak{g}, f_{min})$ is determined by the Lusztig-Feigin-Frenkel conjecture (Lusztig [1991], Arakawa and Fiebig [2012], and E. Frenkel and Gaitsgory [2009]).

Remark 2. Theorem 1 holds in the case that $\mathfrak{g}$ is a basic classical Lie superalgebra as well. In particular one obtains the character of irreducible highest weight representations of superconformal algebras that appear as $\mathcal{W}^k(\mathfrak{g}, f_{min})$ once the character of irreducible highest weight representations of $\widehat{\mathfrak{g}}$ is given.

Let KL_k be the full subcategory of $\mathcal{O}_k$ consisting of objects on which $\mathfrak{g}[t]$ acts locally finitely. Altough the functor

$$\mathcal{O}_k \to \mathcal{W}^k(\mathfrak{g}, f)\text{-Mod}, \quad M \mapsto H^0_{DS,f}(M) \tag{9}$$

is not exact for a general nilpotent element f, we have the following result.

Theorem 3 (Arakawa [2015a]). *Let f, k be arbitrary. We have $H^i_{DS,f_{min}}(M) = 0$ for any $M \in \mathrm{KL}_k$ and $i \neq 0$. Therefore, the functor*

$$\mathrm{KL}_k \to \mathcal{W}^k(\mathfrak{g}, f_{min})\text{-Mod}, \quad M \mapsto H^0_{DS,f_{min}}(M)$$

is exact.

In the case that f is a principal nilpotent element, Theorem 3 has been proved in E. Frenkel and Gaitsgory [2010] using Theorem 4 below.

The restriction of the quantized Drinfeld-Sokolv reduction functor to KL_k does not produce all the irreducible highest weight representations of $\mathcal{W}^k(\mathfrak{g}, f)$. However, one can modify the functor (9) to the "$-$"-reduction functor $H^0_{-,f}(?)$ (defined in E. Frenkel, V. Kac, and Wakimoto [1992]) to obtain the following result for the principal W-algebras.

Theorem 4 (Arakawa [2007])**.** *Let f be a principal nilpotent element, and let k be an arbitrary complex number.*

i. *We have $H^i_{-,f_{prin}}(M) = 0$ for any $M \in \mathcal{O}_k$ and $i \in \mathbb{Z}\backslash\{0\}$. Therefore, the functor*

$$\mathcal{O}_k \to \mathcal{W}^k(\mathfrak{g}, f_{prin})\text{-Mod}, \quad M \mapsto H^0_{-,f_{prin}}(M)$$

is exact.

ii. *For a weight λ of $\widehat{\mathfrak{g}}$ of level k, $H^0_{-,f_{prin}}(L(\lambda))$ is zero or isomorphic to an irreducible highest weight representation of $\mathcal{W}^k(\mathfrak{g}, f_{prin})$. Moreover, any irreducible highest weight representation the principal W-algebra $\mathcal{W}^k(\mathfrak{g}, f_{prin})$ arises in this way.*

In type A we can derive the similar result as Theorem 4 for any nilpotent element f using the work of Brundan and Kleshchev [2008] on the representation theory of finite W-algebras (Arakawa [2011]). In particular the character of all ordinary irreducible representations of $\mathcal{W}^k(\mathfrak{sl}_n, f)$ has been determined for a non-critical k.

5 BRST reduction of associated varieties

Let $\mathcal{W}_k(\mathfrak{g}, f)$ be the unique simple graded quotient of $\mathcal{W}^k(\mathfrak{g}, f)$. The associated variety $X_{\mathcal{W}_k(\mathfrak{g},f)}$ is a subvariety of $X_{\mathcal{W}^k(\mathfrak{g},f)} = \mathcal{S}_f$, which is invariant under the natural $\mathbb{C}^*$-action on $\mathcal{S}_f$ that contracts to the point $f \in \mathcal{S}_f$. Therefore $\mathcal{W}_k(\mathfrak{g}, f)$ is lisse if and only if $X_{\mathcal{W}_k(\mathfrak{g},f)} = \{f\}$.

By Theorem 3, $\mathcal{W}_k(\mathfrak{g}, f)$ is a quotient of the vertex algebra $H^0_{DS,f}(L_k(\mathfrak{g}))$, provided that it is nonzero.

Theorem 5 (Arakawa [2015a])**.** *For any $f \in \mathfrak{g}$ and $k \in \mathbb{C}$ we have*

$$X_{H^0_{DS,f}(L_k(\mathfrak{g}))} \cong X_{L_k(\mathfrak{g})} \cap \mathcal{S}_f.$$

Therefore,

i. *$H^0_{DS,f}(L_k(\mathfrak{g})) \neq 0$ if and only if $\overline{G.f} \subset X_{L_k(\mathfrak{g})}$;*

ii. *If* $X_{L_k(\mathfrak{g})} = \overline{G.f}$ *then* $X_{H^0_{DS,f}(L_k(\mathfrak{g}))} = \{f\}$. *Hence* $H^0_{DS,f}(L_k(\mathfrak{g}))$ *is lisse, and thus, so is its quotient* $\mathcal{W}_k(\mathfrak{g}, f)$.

Theorem 5 can be regarded as a vertex algebra analogue of the corresponding result Losev [2011] and Ginzburg [2009] for finite W-algebras.

Note that if $L_k(\mathfrak{g})$ is integrable we have $H^0_{DS,f}(L_k(\mathfrak{g})) = 0$ by (6). Therefore we need to study more general representations of $\widehat{\mathfrak{g}}$ to obtain lisse W-algebras using Theorem 5.

Recall that the irreducible highest weight representation $L(\lambda)$ of $\widehat{\mathfrak{g}}$ is called *admissible* V. G. Kac and Wakimoto [1989] (1) if λ is regular dominant, that is, $\langle \lambda + \rho, \alpha^\vee \rangle \notin -\mathbb{Z}_{\geq 0}$ for any $\alpha \in \Delta^{re}_+$, and (2) $\mathbb{Q}\Delta(\lambda) = \mathbb{Q}\Delta^{re}$. Here Δ^{re} is the set of real roots of $\widehat{\mathfrak{g}}$, Δ^{re}_+ the set of positive real roots of $\widehat{\mathfrak{g}}$, and $\Delta(\lambda) = \{\alpha \in \Delta^{re} \mid \langle \lambda + \rho, \alpha^\vee \rangle \in \mathbb{Z}\}$, the set of integral roots of λ. Admissible representations are (conjecturally all) modular invariant representations of $\widehat{\mathfrak{g}}$, that is, the characters of admissible representations are invariant under the natural action of $SL_2(\mathbb{Z})$ (V. G. Kac and Wakimoto [1988]). The simple affine vertex algebra $L_k(\mathfrak{g})$ is admissible as a $\widehat{\mathfrak{g}}$-module if and only if

$$k + h^\vee = \frac{p}{q}, \quad p, q \in \mathbb{N},\ (p,q) = 1,\ p \geq \begin{cases} h^\vee & \text{if } (q, r^\vee) = 1, \\ h & \text{if } (q, r^\vee) = r^\vee \end{cases} \tag{10}$$

(V. G. Kac and Wakimoto [2008]). Here h is the Coxeter number of $\mathfrak{g}$ and $r^\vee$ is the lacity of $\mathfrak{g}$. If this is the case k is called an *admissible number* for $\widehat{\mathfrak{g}}$ and $L_k(\mathfrak{g})$ is called an *admissible affine vertex algebra*.

Theorem 6 (Arakawa [2015a])**.** *Let* $L_k(\mathfrak{g})$ *be an admissible affine vertex algebra.*

i. *(Feigin-Frenkel conjectrue) We have* $X_{L_k(\mathfrak{g})} \subset \mathfrak{N}$, *the nilpotent cone of* $\mathfrak{g}$.

ii. *The variety* $X_{L_k(\mathfrak{g})}$ *is irreducible. That is, there exists a nilpotent orbit* $\mathbb{O}_k$ *of* $\mathfrak{g}$ *such that*

$$X_{L_k(\mathfrak{g})} = \overline{\mathbb{O}_k}.$$

iii. *More precisely, let* k *be an admissible number of the form* (10)*. Then*

$$X_{L_k(\mathfrak{g})} = \begin{cases} \{x \in \mathfrak{g} \mid (\operatorname{ad} x)^{2q} = 0\} & \text{if } (q, r^\vee) = 1, \\ \{x \in \mathfrak{g} \mid \pi_{\theta_s}(x)^{2q/r^\vee} = 0\} & \text{if } (q, r^\vee) = r^\vee, \end{cases}$$

where θ_s *is the highest short root of* $\mathfrak{g}$ *and* π_{θ_s} *is the irreducible finite-dimensional representation of* $\mathfrak{g}$ *with highest weight* θ_s.

From Theorem 5 and Theorem 6 we immediately obtain the following assertion, which was (essentially) conjectured by V. G. Kac and Wakimoto [2008].

Theorem 7 (Arakawa [2015a]). *Let $L_k(\mathfrak{g})$ be an admissible affine vertex algebra, and let $f \in \mathbb{O}_k$. Then the simple affine W-algebra $\mathcal{W}_k(\mathfrak{g}, f)$ is lisse.*

In the case that $X_{L_k(\mathfrak{g})} = \overline{G.f_{prin}}$, the lisse W-algebras obtained in Theorem 7 is the *minimal series* principal W-algebras studied in E. Frenkel, V. Kac, and Wakimoto [1992]. In the case that $\mathfrak{g} = \mathfrak{sl}_2$, these are exactly the minimal series Virasoro vertex algebras (Feĭgin and Fuchs [1984], Beilinson, Belavin, V. Drinfeld, and et al. [2004], and Wang [1993]). The Frenkel-Kac-Wakimoto conjecture states that these minimal series principal W-algebras are rational. More generally, all the lisse W-algebras $\mathcal{W}_k(\mathfrak{g}, f)$ that appear in Theorem 7 are conjectured to be rational (V. G. Kac and Wakimoto [2008] and Arakawa [2015a]).

6 The rationality of minimal series principal W-algebras

An admissible affine vertex algebra $L_k(\mathfrak{g})$ is called *non-degenearte* (E. Frenkel, V. Kac, and Wakimoto [1992]) if

$$X_{L_k(\mathfrak{g})} = \mathfrak{N} = \overline{G.f_{prin}}.$$

If this is the case k is called a *non-degenerate admissible number* for $\widehat{\mathfrak{g}}$. By Theorem 6 (iii), "most" admissible affine vertex algebras are non-degenerate. More precisely, an admissible number k of the form (10) is non-degenerate if and only if

$$q \geq \begin{cases} h & \text{if } (q, r^\vee) = 1, \\ r^\vee {}^L h^\vee & \text{if } (q, r^\vee) = r^\vee \end{cases}$$

where ${}^L h^\vee$ is the dual Coxeter number of the Langlands dual Lie algebra ${}^L\mathfrak{g}$. For a non-degenerate admissible number k, the simple principal W-algebra $\mathcal{W}_k(\mathfrak{g}, f_{prin})$ is lisse by Theorem 7.

The following assertion settles the Frenkel-Kac-Waimoto conjecture E. Frenkel, V. Kac, and Wakimoto [ibid.] in full generality.

Theorem 8 (Arakawa [2015b]). *Let k be a non-degenerate admissible number. Then the simple principal W-algebra $\mathcal{W}_k(\mathfrak{g}, f_{prin})$ is rational.*

The proof of Theorem 8 based on Theorem 4, Theorem 7, and the following assertion on admissible affine vertex algebras, which was conjectured by Adamović and Milas [1995].

Theorem 9 (Arakawa [2016]). *Let $L_k(\mathfrak{g})$ be an admissible affine vertex algebra. Then $L_k(\mathfrak{g})$ is rational in the category $\mathcal{O}$, that is, any $L_k(\mathfrak{g})$-module that belongs to $\mathcal{O}$ is completely reducible.*

The following assertion, which has been widely believed since V. G. Kac and Wakimoto [1990] and E. Frenkel, V. Kac, and Wakimoto [1992], gives a yet another realization of minimal series principal W-algebras.

Theorem 10 (Arakawa, Creutzig, and Linshaw [2018]). *Let $\mathfrak{g}$ be simply laced. For an admissible affine vertex algebra $L_k(\mathfrak{g})$, the vertex algebra $(L_k(\mathfrak{g}) \otimes L_1(\mathfrak{g}))^{\mathfrak{g}[t]}$ is isomorphic to a minimal series principal W-algebra. Conversely, any minimal series principal W-algebra associated with $\mathfrak{g}$ appears in this way.*

In the case that $\mathfrak{g} = \mathfrak{sl}_2$ and k is a non-negative integer, the statement of Theorem 10 is well-known as the GKO construction of the discrete series of the Virasoro vertex algebras Goddard, Kent, and Olive [1986]. Some partial results have been obtained previously in Arakawa, Lam, and Yamada [2017] and Arakawa and Jiang [2018]. From Theorem 10, it follows that the minimal series principal W-algebra $\mathcal{W}_{p/q-h^\vee}(\mathfrak{g}, f_{prin})$ of ADE type is unitary, that is, any simple $\mathcal{W}_{p/q-h^\vee}(\mathfrak{g}, f_{prin})$-module is unitary in the sense of Dong and Lin [2014], if and only if $|p-q| = 1$.

7 Four-dimensional $N = 2$ superconformal algebras, Higgs branch conjecture and the class $\mathcal{S}$ chiral algebras

In the study of four-dimensional $N = 2$ superconformal field theories in physics, Beem, Lemos, Liendo, Peelaers, Rastelli, and van Rees Beem, Lemos, Liendo, Peelaers, Rastelli, and van Rees [2015] have constructed a remarkable map

(11) $$\Phi : \{\text{4d } N = 2 \text{ SCFTs}\} \to \{\text{vertex algebras}\}$$

such that, among other things, the character of the vertex algebra $\Phi(\mathcal{T})$ coincides with the *Schur index* of the corresponding 4d $N = 2$ SCFT $\mathcal{T}$, which is an important invariant of the theory $\mathcal{T}$.

How do vertex algebras coming from 4d $N = 2$ SCFTs look like? According to Beem, Lemos, Liendo, Peelaers, Rastelli, and van Rees [ibid.], we have

$$c_{2d} = -12 c_{4d},$$

where c_{4d} and c_{2d} are central charges of the 4d $N = 2$ SCFT and the corresponding vertex algebra, respectively. Since the central charge is positive for a unitary theory, this implies that the vertex algebras obtained by Φ are never unitary. In particular integrable affine vertex algebras never appear by this correspondence.

The main examples of vertex algebras considered in Beem, Lemos, Liendo, Peelaers, Rastelli, and van Rees [ibid.] are the affine vertex algebras $L_k(\mathfrak{g})$ of types D_4, F_4, E_6, E_7,

E_8 at level $k = -h^\vee/6-1$, which are non-rational, non-admissible affine vertex algebras studies in Arakawa and Moreau [2016]. One can find more examples in the literature, see e.g. Beem, Peelaers, Rastelli, and van Rees [2015], Xie, Yan, and Yau [2016], Córdova and Shao [2016], and Song, Xie, and Yan [2017].

Now, there is another important invariant of a 4d $N = 2$ SCFT $\mathcal{T}$, called the *Higgs branch*, which we denote by $Higgs_{\mathcal{T}}$. The Higgs branch $Higgs_{\mathcal{T}}$ is an affine algebraic variety that has the hyperKähler structure in its smooth part. In particular, $Higgs_{\mathcal{T}}$ is a (possibly singular) symplectic variety.

Let $\mathcal{T}$ be one of the 4d $N = 2$ SCFTs studied in Beem, Lemos, Liendo, Peelaers, Rastelli, and van Rees [2015] such that that $\Phi(\mathcal{T}) = L_k(\mathfrak{g})$ with $k = h^\vee/6-1$ for types D_4, F_4, E_6, E_7, E_8 as above. It is known that $Higgs_{\mathcal{T}} = \overline{\mathbb{O}_{min}}$, and this equals Arakawa and Moreau [2016] to the the associated variety $X_{\Phi(\mathcal{T})}$. It is expected that this is not just a coincidence.

Conjecture 11 (Beem and Rastelli [2017])**.** For a 4d $N = 2$ SCFT $\mathcal{T}$, we have

$$Higgs_{\mathcal{T}} = X_{\Phi(\mathcal{T})}.$$

So we are expected to recover the Higgs branch of a 4d $N = 2$ SCFT from the corresponding vertex algebra, which is a purely algebraic object.

We note that Conjecture 11 is a physical conjecture since the Higgs branch is not a mathematical defined object at the moment. The Schur index is not a mathematical defined object either. However, in view of (11) and Conjecture 11, one can try to define both Higgs branches and Schur indices of 4d $N = 2$ SCFTs using vertex algebras. We note that there is a close relationship between Higgs branches of 4d $N = 2$ SCFTs and *Coulomb branches* of three-dimensional $N = 4$ gauge theories whose mathematical definition has been given by Braverman, Finkelberg, and Nakajima [2016b].

Although Higgs branches are symplectic varieties, the associated variety X_V of a vertex algebra V is only a Poisson variety in general. A vertex algebra V is called *quasi-lisse* (Arakawa and Kawasetsu [n.d.]) if X_V has only finitely many symplectic leaves. If this is the case symplectic leaves in X_V are algebraic (Brown and Gordon [2003]). Clearly, lisse vertex algebras are quasi-lisse. The simple affine vertex algebra $L_k(\mathfrak{g})$ is quasi-lisse if and only if $X_{L_k(\mathfrak{g})} \subset \mathfrak{N}$. In particular, admissible affine vertex algebras are quasi-lisse. See Arakawa and Moreau [2016, 2017, 2018b] for more examples of quasi-lisse vertex algebras. Physical intuition expects that vertex algebras that come from 4d $N = 2$ SCFTs via the map Φ are quasi-lisse.

By extending Zhu's argument Zhu [1996] using a theorem of Etingof and Schelder Etingof and Schedler [2010], we obtain the following assertion.

Theorem 12 (Arakawa and Kawasetsu [n.d.])**.** *Let V be a quasi-lisse $\mathbb{Z}_{\geq 0}$-graded conformal vertex algebra such that $V_0 = \mathbb{C}$. Then there only finitely many simple ordinary*

V-modules. Moreover, for an ordinary V-module M, the character $\chi_M(q)$ satisfies a modular linear differential equation.

Since the space of solutions of a modular linear differential equation is invariant under the action of $SL_2(\mathbb{Z})$, Theorem 12 implies that a quasi-lisse vertex algebra possesses a certain modular invariance property, although we do not claim that the normalized characters of ordinary V-modules span the space of the solutions. Note that Theorem 12 implies that the Schur indices of 4d $N = 2$ SCFTs have some modular invariance property. This is something that has been conjectured by physicists (Beem and Rastelli [2017]).

There is a distinct class of four-dimensional $N = 2$ superconformal field theories called the *theory of class* $\mathcal{S}$ (Gaiotto [2012] and Gaiotto, Moore, and Neitzke [2013]), where $\mathcal{S}$ stands for 6. The vertex algebras obtained from the theory of class $\mathcal{S}$ is called the *chiral algebras of class* $\mathcal{S}$ (Beem, Peelaers, Rastelli, and van Rees [2015]). The Moore-Tachikawa conjecture Moore and Tachikawa [2012], which was recently proved in Braverman, Finkelberg, and Nakajima [2017], describes the Higgs branches of the theory of class $\mathcal{S}$ in terms of two-dimensional topological quantum field theories.

Let $\mathbb{V}$ be the category of vertex algebras, whose objects are semisimple groups, and $\mathrm{Hom}(G_1, G_2)$ is the isomorphism classes of conformal vertex algebras V with a vertex algebra homomorphism

$$V^{-h_1^\vee}(\mathfrak{g}_1) \otimes V^{-h_2^\vee}(\mathfrak{g}_2) \to V$$

such that the action of $\mathfrak{g}_1[t] \oplus \mathfrak{g}_2[t]$ on V is locally finite. Here $\mathfrak{g}_i = \mathrm{Lie}(G_i)$ and $h_i^\vee$ is the dual Coxeter number of $\mathfrak{g}_i$ in the case that $\mathfrak{g}_i$ is simple. If $\mathfrak{g}_i$ is not simple we understand $V^{-h_i^\vee}(\mathfrak{g}_i)$ to be the tensor product of the critical level universal affine vertex algebras corresponding to all simple components of $\mathfrak{g}_i$. The composition $V_1 \circ V_2$ of $V_1 \in \mathrm{Hom}(G_1, G_2)$ and $V_2 \in \mathrm{Hom}(G_1, G_2)$ is given by the relative semi-infinite cohomology

$$V_1 \circ V_2 = H^{\frac{\infty}{2}+\bullet}(\widehat{\mathfrak{g}}_2, \mathfrak{g}_2, V_1 \otimes V_2),$$

where $\widehat{\mathfrak{g}}_2$ denotes the direct sum of the affine Kac-Moody algebra associated with the simple components of $\mathfrak{g}_2$. By a result of Arkhipov and Gaitsgory [2002], one finds that the identity morphism id_G is the algebra $\mathfrak{D}_G^{ch}$ of *chiral differential operators* on G (Malikov, Schechtman, and Vaintrob [1999] and Beilinson and V. Drinfeld [2004]) at the critical level, whose associated variety is canonically isomorphic to T^*G.

The following theorem, which was conjectured in Beem, Lemos, Liendo, Peelaers, Rastelli, and van Rees [2015] (see Tachikawa [n.d.(a),(b)]) for mathematical expositions), describes the chiral algebras of class $\mathcal{S}$.

Theorem 13 (Arakawa [n.d.]). *Let $\mathbb{B}_2$ the category of 2-bordisms. There exists a unique monoidal functor $\eta_G : \mathbb{B}_2 \to \mathbb{V}$ which sends (1) the object S^1 to G, (2) the cylinder,* *which is the identity morphism* id_{S^1}, *to the identity morphism* $\mathrm{id}_G = \mathfrak{D}_G^{ch}$, *and (3) the cap to* $H^0_{DS,f_{prin}}(\mathfrak{D}_G^{ch})$. *Moreover, we have* $X_{\eta_G(B)} \cong \eta_G^{BFN}(B)$ *for any 2-bordism* B, *where* η_G^{BFN} *is the functor form* $\mathbb{B}_2$ *to the category of symplectic variaties constructed in Braverman, Finkelberg, and Nakajima [2017].*

The last assertion of the above theorem confirms the Higgs branch conjecture for the theory of class $\mathcal{S}$.

References

Toshiyuki Abe, Geoffrey Buhl, and Chongying Dong (2004). "Rationality, regularity, and C_2-cofiniteness". *Trans. Amer. Math. Soc.* 356.8, pp. 3391–3402. MR: 2052955 (cit. on p. 1285).

Dražen Adamović and Antun Milas (1995). "Vertex operator algebras associated to modular invariant representations for $A_1^{(1)}$". *Math. Res. Lett.* 2.5, pp. 563–575. MR: 1359963 (cit. on p. 1291).

Mina Aganagic, Edward Frenkel, and Andrei Okounkov (Jan. 2017). "Quantum q-Langlands Correspondence". arXiv: 1701.03146 (cit. on p. 1281).

Luis F. Alday, Davide Gaiotto, and Yuji Tachikawa (2010). "Liouville correlation functions from four-dimensional gauge theories". *Lett. Math. Phys.* 91.2, pp. 167–197. MR: 2586871 (cit. on p. 1281).

T. Arakawa and A. Premet (2017). "Quantizing Mishchenko-Fomenko subalgebras for centralizers via affine W-algebras". *Trans. Moscow Math. Soc.* 78, pp. 217–234. MR: 3738086 (cit. on p. 1287).

Tomoyuki Arakawa (n.d.). "Chiral algebras of class $\mathcal{S}$ and symplectic varieties". In preparation (cit. on p. 1294).

– (2005). "Representation theory of superconformal algebras and the Kac-Roan-Wakimoto conjecture". *Duke Math. J.* 130.3, pp. 435–478. MR: 2184567 (cit. on p. 1288).

– (2007). "Representation theory of $\mathcal{W}$-algebras". *Invent. Math.* 169.2, pp. 219–320. MR: 2318558 (cit. on pp. 1286, 1289).

– (2011). "Representation theory of W-algebras, II". In: *Exploring new structures and natural constructions in mathematical physics*. Vol. 61. Adv. Stud. Pure Math. Math. Soc. Japan, Tokyo, pp. 51–90. MR: 2867144 (cit. on p. 1289).

– (2012a). "W-algebras at the critical level". In: *Algebraic groups and quantum groups*. Vol. 565. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 1–13. MR: 2932418 (cit. on p. 1287).

– (2012b). "A remark on the C_2-cofiniteness condition on vertex algebras". *Math. Z.* 270.1-2, pp. 559–575. MR: 2875849 (cit. on pp. 1283, 1284).

Tomoyuki Arakawa (2015a). “Associated varieties of modules over Kac-Moody algebras and C_2-cofiniteness of W-algebras”. *Int. Math. Res. Not. IMRN* 22, pp. 11605–11666. MR: 3456698 (cit. on pp. 1286, 1288–1291).

– (2015b). “Rationality of W-algebras: principal nilpotent cases”. *Ann. of Math. (2)* 182.2, pp. 565–604. MR: 3418525 (cit. on p. 1291).

– (2016). “Rationality of admissible affine vertex algebras in the category $\mathcal{O}$”. *Duke Math. J.* 165.1, pp. 67–93. MR: 3450742 (cit. on p. 1291).

Tomoyuki Arakawa, Thomas Creutzig, and Andrew R Linshaw (2018). “W-algebras as coset vertex algebras”. arXiv: 1801.03822 (cit. on p. 1292).

Tomoyuki Arakawa and Peter Fiebig (2012). “On the restricted Verma modules at the critical level”. *Trans. Amer. Math. Soc.* 364.9, pp. 4683–4712. MR: 2922606 (cit. on p. 1288).

Tomoyuki Arakawa and Cuipo Jiang (2018). “Coset vertex operator algebras and W-algebras of A-type”. *Sci. China Math.* 61.2, pp. 191–206. MR: 3755775 (cit. on p. 1292).

Tomoyuki Arakawa and Kazuya Kawasetsu (n.d.). “Quasi-lisse vertex algebras and modular linear differential equations”. To appear in Kostant Memorial Volume (cit. on p. 1293).

Tomoyuki Arakawa, Ching Hung Lam, and Hiromichi Yamada (Jan. 2017). “Parafermion vertex operator algebras and W-algebras”. To appear in Trans. Amer. Math. Soc. arXiv: 1701.06229 (cit. on p. 1292).

Tomoyuki Arakawa and Alexander Molev (2017). “Explicit generators in rectangular affine $\mathcal{W}$-algebras of type A”. *Lett. Math. Phys.* 107.1, pp. 47–59. MR: 3598875.

Tomoyuki Arakawa and Anne Moreau (2016). “Joseph ideals and lisse minimal W-algebras”. *Journal of the Institute of Mathematics of Jussieu*, pp. 1–21 (cit. on p. 1293).

– (2017). “Sheets and associated varieties of affine vertex algebras”. *Adv. Math.* 320, pp. 157–209. MR: 3709103 (cit. on p. 1293).

– (Feb. 2018a). “Arc spaces and chiral symplectic cores”. arXiv: 1802.06533 (cit. on p. 1287).

– (2018b). “On the irreducibility of associated varieties of W-algebras”. *J. Algebra* 500, pp. 542–568. MR: 3765468 (cit. on p. 1293).

S. Arkhipov and D. Gaitsgory (2002). “Differential operators on the loop group via chiral algebras”. *Int. Math. Res. Not.* 4, pp. 165–210. MR: 1876958 (cit. on p. 1294).

Bojko Bakalov and Todor Milanov (2013). “$\mathcal{W}$-constraints for the total descendant potential of a simple singularity”. *Compos. Math.* 149.5, pp. 840–888. MR: 3069364 (cit. on p. 1281).

Christopher Beem, Madalena Lemos, Pedro Liendo, Wolfger Peelaers, Leonardo Rastelli, and Balt C. van Rees (2015). “Infinite chiral symmetry in four dimensions”. *Comm. Math. Phys.* 336.3, pp. 1359–1433. MR: 3324147 (cit. on pp. 1292–1294).

Christopher Beem, Wolfger Peelaers, Leonardo Rastelli, and Balt C. van Rees (2015). "Chiral algebras of class $\mathcal{S}$". *J. High Energy Phys.* 5, 020, front matter+67. MR: 3359377 (cit. on pp. 1293, 1294).

Christopher Beem and Leonardo Rastelli (July 2017). "Vertex operator algebras, Higgs branches, and modular differential equations". arXiv: 1707.07679 (cit. on pp. 1293, 1294).

A. Beilinson, A. Belavin, V. Drinfeld, and et al. (2004). "Boris Feigin". *Mosc. Math. J.* 4.3, pp. 537–546. MR: 2119138 (cit. on p. 1291).

Alexander Beilinson and Vladimir Drinfeld (2004). *Chiral algebras*. Vol. 51. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, pp. vi+375. MR: 2058353 (cit. on p. 1294).

Richard E. Borcherds (1986). "Vertex algebras, Kac-Moody algebras, and the Monster". *Proc. Nat. Acad. Sci. U.S.A.* 83.10, pp. 3068–3071. MR: 843307 (cit. on p. 1283).

P. Bouwknegt and K. Schoutens, eds. (1995). *$\mathcal{W}$-symmetry*. Vol. 22. Advanced Series in Mathematical Physics. World Scientific Publishing Co., Inc., River Edge, NJ, pp. xxvi+875. MR: 1338864 (cit. on p. 1287).

Alexander Braverman, Michael Finkelberg, and Hiraku Nakajima (2016a). "Instanton moduli spaces and $\mathfrak{W}$-algebras". *Astérisque* 385, pp. vii+128. MR: 3592485 (cit. on p. 1281).

– (Jan. 2016b). "Towards a mathematical definition of Coulomb branches of 3-dimensional $\mathcal{N} = 4$ gauge theories, II". arXiv: 1601.03586 (cit. on p. 1293).

– (June 2017). "Ring objects in the equivariant derived Satake category arising from Coulomb branches (with an appendix by Gus Lonergan)". arXiv: 1706.02112 (cit. on pp. 1294, 1295).

Kenneth A. Brown and Iain Gordon (2003). "Poisson orders, symplectic reflection algebras and representation theory". *J. Reine Angew. Math.* 559, pp. 193–216. MR: 1989650 (cit. on p. 1293).

Jonathan Brundan and Alexander Kleshchev (2008). "Representations of shifted Yangians and finite W-algebras". *Mem. Amer. Math. Soc.* 196.918, pp. viii+107. MR: 2456464 (cit. on p. 1289).

Clay Córdova and Shu-Heng Shao (2016). "Schur indices, BPS particles, and Argyres-Douglas theories". *J. High Energy Phys.* 1, 040, front matter+37. MR: 3471540 (cit. on p. 1293).

Alberto De Sole and Victor G. Kac (2006). "Finite vs affine W-algebras". *Jpn. J. Math.* 1.1, pp. 137–261. MR: 2261064 (cit. on p. 1286).

Alberto De Sole, Victor G. Kac, and Daniele Valeri (2013). "Classical $\mathcal{W}$-algebras and generalized Drinfeld-Sokolov bi-Hamiltonian systems within the theory of Poisson vertex algebras". *Comm. Math. Phys.* 323.2, pp. 663–711. MR: 3096535 (cit. on p. 1281).

Chongying Dong and Xingjun Lin (2014). "Unitary vertex operator algebras". *J. Algebra* 397, pp. 252–277. MR: 3119224 (cit. on p. 1292).

Chongying Dong, Xingjun Lin, and Siu-Hung Ng (2015). "Congruence property in conformal field theory". *Algebra Number Theory* 9.9, pp. 2121–2166. MR: 3435813 (cit. on p. 1285).

Chongying Dong and Geoffrey Mason (2006). "Integrability of C_2-cofinite vertex operator algebras". *Int. Math. Res. Not.* Art. ID 80468, 15. MR: 2219226 (cit. on p. 1284).

V. G. Drinfeld and V. V. Sokolov (1984). "Lie algebras and equations of Korteweg-de Vries type". In: *Current problems in mathematics, Vol. 24*. Itogi Nauki i Tekhniki. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, pp. 81–180. MR: 760998 (cit. on p. 1281).

Pavel Etingof and Travis Schedler (2010). "Poisson traces and D-modules on Poisson varieties". *Geom. Funct. Anal.* 20.4. With an appendix by Ivan Losev, pp. 958–987. MR: 2729282 (cit. on p. 1293).

V. A. Fateev and S. L. Lykyanov (1988). "The models of two-dimensional conformal quantum field theory with Z_n symmetry". *Internat. J. Modern Phys. A* 3.2, pp. 507–520. MR: 932661 (cit. on p. 1287).

B. L. Feĭgin (1984). "Semi-infinite homology of Lie, Kac-Moody and Virasoro algebras". *Uspekhi Mat. Nauk* 39.2(236), pp. 195–196. MR: 740035 (cit. on p. 1286).

B. L. Feĭgin and D. B. Fuchs (1984). "Verma modules over the Virasoro algebra". In: *Topology (Leningrad, 1982)*. Vol. 1060. Lecture Notes in Math. Springer, Berlin, pp. 230–245. MR: 770243 (cit. on p. 1291).

Boris Feigin and Edward Frenkel (1990). "Quantization of the Drinfeld-Sokolov reduction". *Phys. Lett. B* 246.1-2, pp. 75–81. MR: 1071340 (cit. on pp. 1286, 1287).

– (1992). "Affine Kac-Moody algebras at the critical level and Gelfand-Dikiĭ algebras". In: *Infinite analysis, Part A, B (Kyoto, 1991)*. Vol. 16. Adv. Ser. Math. Phys. World Sci. Publ., River Edge, NJ, pp. 197–215. MR: 1187549 (cit. on p. 1287).

Edward Frenkel (2007). *Langlands correspondence for loop groups*. Vol. 103. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, pp. xvi+379. MR: 2332156 (cit. on p. 1281).

Edward Frenkel and David Ben-Zvi (2004). *Vertex algebras and algebraic curves*. Second. Vol. 88. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, pp. xiv+400. MR: 2082709 (cit. on pp. 1282, 1283, 1286, 1287).

Edward Frenkel and Dennis Gaitsgory (2009). "Localization of $\mathfrak{g}$-modules on the affine Grassmannian". *Ann. of Math. (2)* 170.3, pp. 1339–1381. MR: 2600875 (cit. on p. 1288).

– (2010). "Weyl modules and opers without monodromy". In: *Arithmetic and geometry around quantization*. Vol. 279. Progr. Math. Birkhäuser Boston, Inc., Boston, MA, pp. 101–121. MR: 2656944 (cit. on p. 1288).

Edward Frenkel, Victor Kac, and Minoru Wakimoto (1992). "Characters and fusion rules for W-algebras via quantized Drinfeld-Sokolov reduction". *Comm. Math. Phys.* 147.2, pp. 295–328. MR: 1174415 (cit. on pp. 1281, 1289, 1291, 1292).

Igor B. Frenkel and Yongchang Zhu (1992). "Vertex operator algebras associated to representations of affine and Virasoro algebras". *Duke Math. J.* 66.1, pp. 123–168. MR: 1159433 (cit. on p. 1285).

Davide Gaiotto (2012). "$N = 2$ dualities". *J. High Energy Phys.* 8, 034, front matter + 57. MR: 3006961 (cit. on p. 1294).

Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke (2013). "Wall-crossing, Hitchin systems, and the WKB approximation". *Adv. Math.* 234, pp. 239–403. MR: 3003931 (cit. on p. 1294).

Dennis Gaitsgory (Jan. 2016). "Quantum Langlands Correspondence". arXiv: 1601.05279 (cit. on p. 1281).

Wee Liang Gan and Victor Ginzburg (2002). "Quantization of Slodowy slices". *Int. Math. Res. Not.* 5, pp. 243–255. MR: 1876934 (cit. on pp. 1286, 1287).

Victor Ginzburg (2009). "Harish-Chandra bimodules for quantized Slodowy slices". *Represent. Theory* 13, pp. 236–271. MR: 2515934 (cit. on p. 1290).

P. Goddard, A. Kent, and D. Olive (1986). "Unitary representations of the Virasoro and super-Virasoro algebras". *Comm. Math. Phys.* 103.1, pp. 105–119. MR: 826859 (cit. on p. 1292).

Yi-Zhi Huang (2008). "Rigidity and modularity of vertex tensor categories". *Commun. Contemp. Math.* 10.suppl. 1, pp. 871–911. MR: 2468370 (cit. on p. 1285).

V. G. Kac and M. Wakimoto (1989). "Classification of modular invariant representations of affine algebras". In: *Infinite-dimensional Lie algebras and groups (Luminy-Marseille, 1988)*. Vol. 7. Adv. Ser. Math. Phys. World Sci. Publ., Teaneck, NJ, pp. 138–177. MR: 1026952 (cit. on p. 1290).

– (1990). "Branching functions for winding subalgebras and tensor products". *Acta Appl. Math.* 21.1-2, pp. 3–39. MR: 1085771 (cit. on p. 1292).

Victor Kac (1998). *Vertex algebras for beginners*. Second. Vol. 10. University Lecture Series. American Mathematical Society, Providence, RI, pp. vi+201. MR: 1651389 (cit. on p. 1282).

– (Dec. 2015). "Introduction to vertex algebras, Poisson vertex algebras, and integrable Hamiltonian PDE". To appear in *Perspectives in Lie Theory*, Springer INdAM Series 19. arXiv: 1512.00821 (cit. on p. 1283).

Victor G. Kac and Minoru Wakimoto (1988). "Modular invariant representations of infinite-dimensional Lie algebras and superalgebras". *Proc. Nat. Acad. Sci. U.S.A.* 85.14, pp. 4956–4960. MR: 949675 (cit. on p. 1290).

– (2004). "Quantum reduction and representation theory of superconformal algebras". *Adv. Math.* 185.2, pp. 400–458. MR: 2060475 (cit. on p. 1286).

Victor G. Kac and Minoru Wakimoto (2008). "On rationality of W-algebras". *Transform. Groups* 13.3-4, pp. 671–713. MR: 2452611 (cit. on pp. 1290, 1291).

Victor Kac, Shi-Shyr Roan, and Minoru Wakimoto (2003). "Quantum reduction for affine superalgebras". *Comm. Math. Phys.* 241.2-3, pp. 307–342. MR: 2013802 (cit. on pp. 1286, 1287).

Masaki Kashiwara and Toshiyuki Tanisaki (2000). "Characters of irreducible modules with non-critical highest weights over affine Lie algebras". In: *Representations and quantizations (Shanghai, 1998)*. China High. Educ. Press, Beijing, pp. 275–296. MR: 1802178 (cit. on p. 1288).

Haisheng Li (2005). "Abelianizing vertex algebras". *Comm. Math. Phys.* 259.2, pp. 391–411. MR: 2172688 (cit. on pp. 1283, 1284).

Ivan Losev (2011). "Finite-dimensional representations of W-algebras". *Duke Math. J.* 159.1, pp. 99–143. MR: 2817650 (cit. on p. 1290).

George Lusztig (1991). "Intersection cohomology methods in representation theory". In: *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*. Math. Soc. Japan, Tokyo, pp. 155–174. MR: 1159211 (cit. on p. 1288).

Fyodor Malikov, Vadim Schechtman, and Arkady Vaintrob (1999). "Chiral de Rham complex". *Comm. Math. Phys.* 204.2, pp. 439–473. MR: 1704283 (cit. on p. 1294).

Davesh Maulik and Andrei Okounkov (Nov. 2012). "Quantum Groups and Quantum Cohomology". arXiv: 1211.1287 (cit. on p. 1281).

Gregory W. Moore and Yuji Tachikawa (2012). "On 2d TQFTs whose values are holomorphic symplectic varieties". In: *String-Math 2011*. Vol. 85. Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI, pp. 191–207. MR: 2985331 (cit. on p. 1294).

Alexander Premet (2002). "Special transverse slices and their enveloping algebras". *Adv. Math.* 170.1. With an appendix by Serge Skryabin, pp. 1–55. MR: 1929302 (cit. on pp. 1281, 1286, 1287).

O. Schiffmann and E. Vasserot (2013). "Cherednik algebras, W-algebras and the equivariant cohomology of the moduli space of instantons on $\mathbf{A}^2$". *Publ. Math. Inst. Hautes Études Sci.* 118, pp. 213–342. MR: 3150250 (cit. on p. 1281).

Jaewon Song, Dan Xie, and Wenbin Yan (June 2017). "Vertex operator algebras of Argyres-Douglas theories from M5-branes". arXiv: 1706.01607 (cit. on p. 1293).

Yuji Tachikawa (n.d.[a]). *On some conjectures on VOAs* (cit. on p. 1294).

– (n.d.[b]). *On'categories' of quantum field theories*. In this volume (cit. on p. 1294).

Meng-Chwan Tan (July 2016). "Higher AGT Correspondences, W-algebras, and Higher Quantum Geometric Langlands Duality from M-Theory". arXiv: 1607.08330 (cit. on p. 1281).

Weiqiang Wang (1993). "Rationality of Virasoro vertex operator algebras". *Internat. Math. Res. Notices* 7, pp. 197–211. MR: 1230296 (cit. on p. 1291).

Dan Xie, Wenbin Yan, and Shing-Tung Yau (Apr. 2016). "Chiral algebra of Argyres-Douglas theory from M5 brane". arXiv: 1604.02155 (cit. on p. 1293).

Yongchang Zhu (1996). "Modular invariance of characters of vertex operator algebras". *J. Amer. Math. Soc.* 9.1, pp. 237–302. MR: 1317233 (cit. on pp. 1285, 1293).

Received 2017-12-06.

Tomoyuki Arakawa (荒川知幸)
arakawa@kurims.kyoto-u.ac.jp

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1301–1320)

DOUBLE AFFINE GRASSMANNIANS AND COULOMB BRANCHES OF $3d\,\mathcal{N}=4$ QUIVER GAUGE THEORIES

Michael Finkelberg

Abstract

We propose a conjectural construction of various slices for double affine Grassmannians as Coulomb branches of 3-dimensional $\mathcal{N}=4$ supersymmetric affine quiver gauge theories. It generalizes the known construction for the usual affine Grassmannians, and makes sense for arbitrary symmetric Kac-Mody algebras.

1 Introduction

1.1 Historical background. The geometric Satake equivalence Lusztig [1983], Ginzburg [1995], Beilinson and Drinfeld [2000], and Mirković and Vilonen [2007] proposed by V. Drinfeld for the needs of the Geometric Langlands Program proved very useful for the study of representation theory of reductive algebraic groups (starting from G. Lusztig's construction of q-analogues of weight multiplicities). About 15 years ago, I. Frenkel and I. Grojnowski envisioned an extension of the geometric Satake equivalence to the case of loop groups. The affine Grassmannians (the main objects of the geometric Satake equivalence) are ind-schemes of ind-finite type. Their loop analogues (double affine Grassmannians) are much more infinite, beyond our current technical abilities. We are bound to settle for some provisional substitutes, such as transversal slices to the smaller strata in the closures of bigger strata. These substitutes still carry quite powerful geometric information.

Following I. Frenkel's suggestion, some particular slices for the double affine Grassmannians were constructed in terms of Uhlenbeck compactifications of instanton moduli spaces on Kleinian singularities about 10 years ago. More recently, H. Nakajima's approach to Coulomb branches of 3-dimensional $\mathcal{N}=4$ supersymmetric gauge theories, applied to affine quiver gauge theories, paved a way for the construction of the most general slices.

MSC2010: primary 22E46; secondary 14J60, 14J81.

1.2 Contents. We recall the geometric Satake equivalence in Section 2. The (generalized) slices for the affine Grassmannians are reviewed in Section 3. The problem of constructing (the slices for) the double affine Grassmannians is formulated in Section 4. The mathematical construction of Coulomb branches of $3d$ $\mathcal{N}=4$ gauge theories and its application to slices occupies Section 5. Some more applications are mentioned in Section 6.

Acknowledgments. This report is based mostly on the works of A. Braverman and H. Nakajima, some joint with the author. I was incredibly lucky to have an opportunity to learn mathematics from them. Before meeting them, I was introduced to some semiinfinite ideas sketched below by A. Beilinson, V. Drinfeld, B. Feigin, V. Ginzburg and I. Mirković. It is also a pleasure to acknowledge my intellectual debt to R. Bezrukavnikov, D. Gaiotto, D. Gaitsgory, J. Kamnitzer and V. Pestun.

2 Geometric Satake equivalence

Let $\mathcal{O}$ denote the formal power series ring $\mathbb{C}[[z]]$, and let $\mathcal{K}$ denote its fraction field $\mathbb{C}((z))$. Let G be an almost simple complex algebraic group with a Borel and a Cartan subgroup $G \supset B \supset T$, and with the Weyl group W_{fin} of (G,T). Let Λ be the coweight lattice, and let $\Lambda^+ \subset \Lambda$ be the submonoid of dominant coweights. Let also $\Lambda_+ \subset \Lambda$ be the submonoid spanned by the simple coroots α_i, $i \in I$. We denote by $G^\vee \supset T^\vee$ the Langlands dual group, so that Λ is the weight lattice of $G^\vee$.

The affine Grassmannian $\mathrm{Gr}_G = G_{\mathcal{K}}/G_{\mathcal{O}}$ is an ind-projective scheme, the union $\bigsqcup_{\bar\lambda\in\Lambda^+}\mathrm{Gr}_G^{\bar\lambda}$ of $G_{\mathcal{O}}$-orbits. The closure of $\mathrm{Gr}_G^{\bar\lambda}$ is a projective variety $\overline{\mathrm{Gr}}_G^{\bar\lambda} = \bigsqcup_{\bar\mu\le\bar\lambda}\mathrm{Gr}_G^{\bar\mu}$. The fixed point set Gr_G^T is naturally identified with the coweight lattice Λ; and $\bar\mu\in\Lambda$ lies in $\mathrm{Gr}_G^{\bar\lambda}$ iff $\bar\mu \in W_{\text{fin}}\bar\lambda$.

One of the cornerstones of the Geometric Langlands Program initiated by V. Drinfeld is an equivalence $\mathbb{S}$ of the tensor category $\mathrm{Rep}(G^\vee)$ and the category $\mathrm{Perv}_{G_{\mathcal{O}}}(\mathrm{Gr}_G)$ of $G_{\mathcal{O}}$-equivariant perverse constructible sheaves on Gr_G equipped with a natural monoidal convolution structure $\star$ and a fiber functor $H^\bullet(\mathrm{Gr}_G,-)$ Lusztig [1983], Ginzburg [1995], Beilinson and Drinfeld [2000], and Mirković and Vilonen [2007]. It is a categorification of the classical Satake isomorphism between $K(\mathrm{Rep}(G^\vee)) = \mathbb{C}[T^\vee]^{W_{\text{fin}}}$ and the spherical affine Hecke algebra of G. The geometric Satake equivalence $\mathbb{S}$ sends an irreducible $G^\vee$-module $V^{\bar\lambda}$ with highest weight $\bar\lambda$ to the Goresky-MacPherson sheaf $\mathrm{IC}(\overline{\mathrm{Gr}}_G^{\bar\lambda})$.

In order to construct a commutativity constraint for $(\mathrm{Perv}_{G_{\mathcal{O}}}(\mathrm{Gr}_G),\star)$, Beilinson and Drinfeld introduced a relative version $\mathrm{Gr}_{G,BD}$ of the Grassmannian over the Ran space of a smooth curve X, and a fusion monoidal structure Ψ on $\mathrm{Perv}_{G_{\mathcal{O}}}(\mathrm{Gr}_G)$ (isomorphic to $\star$). One of the main discoveries of Mirković and Vilonen [2007] was a Λ-grading of the fiber functor $H^\bullet(\mathrm{Gr}_G, \mathcal{F}) = \bigoplus_{\bar\lambda\in\Lambda} \Phi_{\bar\lambda}(\mathcal{F})$ by the hyperbolic stalks at T-fixed points. For a $G^\vee$-module V, its weight space $V_{\bar\lambda}$ is canonically isomorphic to the hyperbolic stalk $\Phi_{\bar\lambda}(\mathbb{S}V)$.

Various geometric structures of a perverse sheaf $\mathbb{S}V$ reflect some fine representation theoretic structures of V, such as Brylinski-Kostant filtration and the action of dynamical Weyl group, see Ginzburg and Riche [2015]. One of the important technical tools of studying $\mathrm{Perv}_{G_\mathcal{O}}(\mathrm{Gr}_G)$ is the embedding $\mathrm{Gr}_G \hookrightarrow \mathbf{Gr}_G$ into Kashiwara infinite type scheme $\mathbf{Gr}_G = G_{\mathbb{C}((z^{-1}))}/G_{\mathbb{C}[z]}$ Kashiwara [1989] and Kashiwara and Tanisaki [1995]. The quotient $G_{\mathbb{C}[[z^{-1}]]}\backslash\mathbf{Gr}_G$ is the moduli stack $\mathrm{Bun}_G(\mathbb{P}^1)$ of G-bundles on the projective line $\mathbb{P}^1$. The $G_{\mathbb{C}[[z^{-1}]]}$-orbits on $\mathbf{Gr}_G$ are of finite codimension; they are also numbered by the dominant coweights of G, and the image of an orbit $\mathbf{Gr}_G^{\bar\lambda}$ in $\mathrm{Bun}_G(\mathbb{P}^1)$ consists of G-bundles of isomorphism type $\bar\lambda$ Grothendieck [1957]. The stratifications $\mathrm{Gr}_G = \bigsqcup_{\bar\lambda\in\Lambda^+} \mathrm{Gr}_G^{\bar\lambda}$ and $\mathbf{Gr}_G = \bigsqcup_{\bar\lambda\in\Lambda^+} \mathbf{Gr}_G^{\bar\lambda}$ are transversal, and their intersections and various generalizations thereof are the subject of the next section.

3 Generalized slices

3.1 The dominant case. We denote by K_1 the first congruence subgroup of $G_{\mathbb{C}[[z^{-1}]]}$: the kernel of the evaluation projection $\mathrm{ev}_\infty\colon G_{\mathbb{C}[[z^{-1}]]} \twoheadrightarrow G$. The transversal slice $\mathcal{W}_{\bar\mu}^{\bar\lambda}$ (resp. $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$) is defined as the intersection of $\mathrm{Gr}_G^{\bar\lambda}$ (resp. $\overline{\mathrm{Gr}}_G^{\bar\lambda}$) and $K_1\cdot\bar\mu$ in $\mathbf{Gr}_G$. It is known that $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$ is nonempty iff $\bar\mu \le \bar\lambda$, and $\dim \overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$ is an affine irreducible variety of dimension $\langle 2\bar\rho^\vee, \bar\lambda - \bar\mu\rangle$. Following an idea of I. Mirković, Kamnitzer, Webster, Weekes, and Yacobi [2014] proved that $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda} = \bigsqcup_{\bar\mu\le\bar\nu\le\bar\lambda} \mathcal{W}_{\bar\mu}^{\bar\nu}$ is the decomposition of $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$ into symplectic leaves of a natural Poisson structure.

The only T-fixed point of $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$ is $\bar\mu$. We consider the cocharacter $2\bar\rho\colon \mathbb{C}^\times \to T$, and denote by $R_{\bar\mu}^{\bar\lambda} \subset \overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$ the corresponding repellent: the closed affine subvariety formed by all the points that flow into $\bar\mu$ under the action of $2\bar\rho(t)$, as t goes to ∞. Let r stand for the closed embedding of $R_{\bar\mu}^{\bar\lambda}$ into $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$, and let ι stand for the closed embedding of $\bar\mu$ into $R_{\bar\mu}^{\bar\lambda}$. Then the hyperbolic stalk $\Phi_{\bar\mu}^{\bar\lambda}\mathcal{F}$ of a T-equivariant constructible complex $\mathcal{F}$ on $\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda}$ is defined as $\iota^! r^*\mathcal{F}$, see Braden [2003] and Drinfeld and Gaitsgory [2014].

Recall that the geometric Satake equivalence takes an irreducible $G^\vee$-module $V^{\bar\lambda}$ to the IC-sheaf $\mathrm{IC}(\overline{\mathrm{Gr}}_G^{\bar\lambda})$, and the weight space $V_{\bar\mu}^{\bar\lambda}$ is realized as $V_{\bar\mu}^{\bar\lambda} = \Phi_{\bar\mu}\mathrm{IC}(\overline{\mathrm{Gr}}_G^{\bar\lambda}) = \Phi_{\bar\mu}^{\bar\lambda}\mathrm{IC}(\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda})$. The usual stalks of both $\mathrm{IC}(\overline{\mathrm{Gr}}_G^{\bar\lambda})$ and $\mathrm{IC}(\overline{\mathcal{W}}_{\bar\mu}^{\bar\lambda})$ at $\bar\mu$ are isomorphic up to shift to the associated graded $\mathrm{gr}\, V_{\bar\mu}^{\bar\lambda}$ with respect to the Brylinski-Kostant filtration.

3.2 The general case. If we want to reconstruct the whole of $V^{\bar{\lambda}}$ from the various slices $\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}}$, we are missing the weight spaces $V^{\bar{\lambda}}_{\bar{\mu}}$ with nondominant $\bar{\mu}$. To take care of the remaining weight spaces, for arbitrary $\bar{\mu}$ we consider the moduli space $\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}}$ of the following data:

(a) A G-bundle $\mathcal{P}$ on $\mathbb{P}^1$.

(b) A trivialization $\sigma\colon \mathcal{P}_{\mathrm{triv}}|_{\mathbb{P}^1\setminus\{0\}} \xrightarrow{\sim} \mathcal{P}|_{\mathbb{P}^1\setminus\{0\}}$ having a pole of degree $\leq \bar{\lambda}$ at $0 \in \mathbb{P}^1$ (that is defining a point of $\overline{\mathrm{Gr}}^{\bar{\lambda}}_G$).

(c) A B-structure ϕ on $\mathcal{P}$ of degree $w_0\bar{\mu}$ with the fiber $B_- \subset G$ at $\infty \in \mathbb{P}^1$ (with respect to the trivialization σ of $\mathcal{P}$ at $\infty \in \mathbb{P}^1$). Here $G \supset B_- \supset T$ is the Borel subgroup opposite to B, and $w_0 \in W_{\mathrm{fin}}$ is the longest element.

This construction goes back to Finkelberg and Mirković [1999]. The space $\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}}$ is nonempty iff $\bar{\mu} \leq \bar{\lambda}$. In this case it is an irreducible affine normal Cohen-Macaulay variety of dimension $\langle 2\bar{\rho}^\vee, \bar{\lambda} - \bar{\mu}\rangle$, see Braverman, Finkelberg, and Nakajima [2016a]. In case $\bar{\mu}$ is dominant, the two definitions of $\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}}$ agree. At the other extreme, if $\bar{\lambda} = 0$, then $\overline{\mathcal{W}}^{0}_{-\alpha}$ is nothing but the open zastava space $\mathring{Z}^{-w_0\alpha}$. The T-fixed point set $(\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}})^T$ is nonempty iff the weight space $V^{\bar{\lambda}}_{\bar{\mu}}$ is not 0; in this case $(\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}})^T$ consists of a single point denoted $\bar{\mu}$. We consider the repellent $R^{\bar{\lambda}}_{\bar{\mu}} \subset \overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}}$. It is a closed subvariety of dimension $\langle \bar{\rho}^\vee, \bar{\lambda} - \bar{\mu}\rangle$ (equidimensional). We have $V^{\bar{\lambda}}_{\bar{\mu}} = \Phi_{\bar{\mu}} \mathrm{IC}(\overline{\mathrm{Gr}}^{\bar{\lambda}}_G) = \Phi^{\bar{\lambda}}_{\bar{\mu}} \mathrm{IC}(\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}})$, so that $V^{\bar{\lambda}} = \bigoplus_{\bar{\mu}\in\Lambda} \Phi^{\bar{\lambda}}_{\bar{\mu}} \mathrm{IC}(\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}})$ (see Krylov [2017]). Similarly to Braverman and Gaitsgory [2001], one can introduce a crystal structure on the set of irreducible components $\bigsqcup_{\bar{\mu}\in\Lambda} \mathrm{Irr}\, R^{\bar{\lambda}}_{\bar{\mu}}$ (see Krylov [2017]), so that the resulting crystal is isomorphic to the integrable crystal $\mathbf{B}(\bar{\lambda})$ (for a beautiful survey on crystals, see Kashiwara [1995]).

3.3 Beilinson-Drinfeld slices. Let $\underline{\bar{\lambda}} = (\bar{\lambda}_1, \ldots, \bar{\lambda}_N)$ be a collection of dominant coweights of G. We consider the moduli space $\underline{\overline{\mathcal{W}}}^{\underline{\bar{\lambda}}}_{\bar{\mu}}$ of the following data:

(a) A collection of points $(z_1, \ldots, z_N) \in \mathbb{A}^N$ on the affine line $\mathbb{A}^1 \subset \mathbb{P}^1$.

(b) A G-bundle $\mathcal{P}$ on $\mathbb{P}^1$.

(c) A trivialization $\sigma\colon \mathcal{P}_{\mathrm{triv}}|_{\mathbb{P}^1\setminus\{z_1,\ldots,z_N\}} \xrightarrow{\sim} \mathcal{P}|_{\mathbb{P}^1\setminus\{z_1,\ldots,z_N\}}$ with a pole of degree $\leq \sum_{s=1}^N \bar{\lambda}_s \cdot z_s$ on the complement.

(d) A B-structure ϕ on $\mathcal{P}$ of degree $w_0\bar{\mu}$ with the fiber $B_- \subset G$ at $\infty \in \mathbb{P}^1$ (with respect to the trivialization σ of $\mathcal{P}$ at $\infty \in \mathbb{P}^1$).

$\underline{\overline{\mathcal{W}}}^{\underline{\bar{\lambda}}}_{\bar{\mu}}$ is nonempty iff $\bar{\mu} \leq \bar{\lambda} := \sum_{s=1}^N \bar{\lambda}_s$. In this case it is an irreducible affine normal Cohen-Macaulay variety flat over $\mathbb{A}^N$ of relative dimension $\langle 2\bar{\rho}^\vee, \bar{\lambda} - \bar{\mu}\rangle$, see Braverman, Finkelberg, and Nakajima [2016a]. The fiber over $N \cdot 0 \in \mathbb{A}^N$ is nothing but $\overline{\mathcal{W}}^{\bar{\lambda}}_{\bar{\mu}}$. We can consider the Verdier specialization $\mathrm{Sp}\,\mathrm{IC}(\underline{\overline{\mathcal{W}}}^{\bar\lambda}_{\bar\mu})$ to the special fiber $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$. It is a perverse sheaf on $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}\times\mathbb{A}^N$ smooth along the diagonal stratification of $\mathbb{A}^N$. We denote by $\Psi\mathrm{IC}(\underline{\overline{\mathcal{W}}}^{\bar\lambda}_{\bar\mu})$ its restriction to $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}\times\underline{z}$ where $\underline{z}$ is a point of $\mathbb{A}^N_{\mathbb{R}}$ such that $z_1>\ldots>z_N$. Then

$$\Psi\mathrm{IC}(\underline{\overline{\mathcal{W}}}^{\bar\lambda}_{\bar\mu})\simeq\bigoplus_{\bar\mu\le\bar\nu\le\bar\lambda,\ \bar\nu\in\Lambda^+}M^{\underline{\bar\lambda}}_{\bar\nu}\otimes\mathrm{IC}(\overline{\mathcal{W}}^{\bar\nu}_{\bar\mu}),$$

where $M^{\underline{\bar\lambda}}_{\bar\nu}$ is the multiplicity $\mathrm{Hom}_{G^\vee}(V^{\bar\nu},V^{\bar\lambda_1}\otimes\ldots\otimes V^{\bar\lambda_N})$.

3.4 Convolution diagram over slices. In the setup of Section 3.3 we consider the moduli space $\underline{\widetilde{\mathcal{W}}}^{\bar\lambda}_{\bar\mu}$ of the following data:

(a) A collection of points $(z_1,\ldots,z_N)\in\mathbb{A}^N$ on the affine line $\mathbb{A}^1\subset\mathbb{P}^1$.

(b) A collection of G-bundles $(\mathcal{P}_1,\ldots,\mathcal{P}_N)$ on $\mathbb{P}^1$.

(c) A collection of isomorphisms $\sigma_s\colon\mathcal{P}_{s-1}|_{\mathbb{P}^1\setminus\{z_s\}}\xrightarrow{\sim}\mathcal{P}_s|_{\mathbb{P}^1\setminus\{z_s\}}$ with a pole of degree $\le\bar\lambda_s$ at z_s. Here $1\le s\le N$, and $\mathcal{P}_0:=\mathcal{P}_{\mathrm{triv}}$.

(d) A B-structure ϕ on $\mathcal{P}_N$ of degree $w_0\bar\mu$ with the fiber $B_-\subset G$ at $\infty\in\mathbb{P}^1$ (with respect to the trivialization $\sigma_N\circ\ldots\circ\sigma_1$ of $\mathcal{P}_N$ at $\infty\in\mathbb{P}^1$).

A natural projection $\varpi\colon\underline{\widetilde{\mathcal{W}}}^{\bar\lambda}_{\bar\mu}\to\underline{\overline{\mathcal{W}}}^{\bar\lambda}_{\bar\mu}$ sends $(\mathcal{P}_1,\ldots,\mathcal{P}_N,\sigma_1,\ldots,\sigma_N)$ to $(\mathcal{P}_N,\sigma_N\circ\ldots\circ\sigma_1)$. We denote $\varpi^{-1}(\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu})$ by $\widetilde{\mathcal{W}}^{\underline{\bar\lambda}}_{\bar\mu}$. Then $\varpi\colon\widetilde{\mathcal{W}}^{\underline{\bar\lambda}}_{\bar\mu}\to\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ is stratified semismall, and

$$\varpi_*\mathrm{IC}(\widetilde{\mathcal{W}}^{\underline{\bar\lambda}}_{\bar\mu})=\bigoplus_{\bar\mu\le\bar\nu\le\bar\lambda,\ \bar\nu\in\Lambda^+}M^{\underline{\bar\lambda}}_{\bar\nu}\otimes\mathrm{IC}(\overline{\mathcal{W}}^{\bar\nu}_{\bar\mu}).$$

4 Double affine Grassmannian

In this section G is assumed to be a simply connected almost simple complex algebraic group.

4.1 The affine group and its Langlands dual. We consider the minimal integral even positive definite W_{fin}-invariant symmetric bilinear form $(\cdot,\cdot)$ on the coweight lattice Λ. It gives rise to a central extension $\widehat{G}$ of the polynomial version $G_{\mathbb{C}[t^{\pm1}]}$ of the loop group:

$$1\to\mathbb{C}^\times\to\widehat{G}\to G_{\mathbb{C}[t^{\pm1}]}\to 1.$$

The loop rotation group $\mathbb{C}^\times$ acts naturally on $G_{\mathbb{C}[t^{\pm1}]}$, and this action lifts to $\widehat{G}$. We denote the corresponding semidirect product $\mathbb{C}^\times\ltimes\widehat{G}$ by G_{aff}. It is an untwisted affine Kac-Moody group ind-scheme.

We denote by $G_{\mathrm{aff}}^{\vee}$ the corresponding Langlands dual group. Note that if G is not simply laced, then $G_{\mathrm{aff}}^{\vee}$ is a twisted affine Kac-Moody group, not to be confused with $(G^{\vee})_{\mathrm{aff}}$. However, we have a canonical embedding $G^{\vee} \hookrightarrow G_{\mathrm{aff}}^{\vee}$.

We fix a Cartan torus $\mathbb{C}^{\times} \times T \times \mathbb{C}^{\times} \subset G_{\mathrm{aff}}$ and its dual Cartan torus $\mathbb{C}^{\times} \times T^{\vee} \times \mathbb{C}^{\times} \subset G_{\mathrm{aff}}^{\vee}$. Here the first copy of $\mathbb{C}^{\times}$ is the central $\mathbb{C}^{\times}$, while the second copy is the loop rotation $\mathbb{C}^{\times}$. Accordingly, the weight lattice Λ_{aff} of $G_{\mathrm{aff}}^{\vee}$ is $\mathbb{Z} \oplus \Lambda \oplus \mathbb{Z}$: the first copy of $\mathbb{Z}$ is the central charge (level), and the second copy is the energy. A typical element $\lambda \in \Lambda_{\mathrm{aff}}$ will be written as $\lambda = (k, \bar{\lambda}, n)$. The subset of dominant weights $\Lambda_{\mathrm{aff}}^{+} \subset \Lambda_{\mathrm{aff}}$ consists of all the triples $(k, \bar{\lambda}, n)$ such that $\bar{\lambda} \in \Lambda^{+}$ and $\langle \bar{\lambda}, \bar{\theta}^{\vee} \rangle \leq k$. Here $\bar{\theta}^{\vee} = \sum_{i \in I} a_i \alpha_i^{\vee}$ is the highest root of $G \supset B \supset T$. We denote by $\Lambda_{\mathrm{aff},k}^{+} \subset \Lambda_{\mathrm{aff}}^{+}$ the finite subset of dominant weights of level k; we also denote by $\Lambda_{\mathrm{aff},k} \subset \Lambda_{\mathrm{aff}}$ the subset of all the weights of level k. We say that $\lambda \geq \mu$ if $\lambda - \mu$ is an element of the submonoid generated by the positive roots of $G_{\mathrm{aff}}^{\vee}$ (in particular, λ and μ must have the same level). Finally, let $\bar{\omega}_i$, $i \in I$, be the fundamental coweights of G, and $\rho := (1, 0, 0) + \sum_{i \in I} (a_i, \bar{\omega}_i, 0) \in \Lambda_{\mathrm{aff}}$.

The affine Weyl group W_{aff} is the semidirect product $W_{\mathrm{fin}} \ltimes \Lambda$. For $k \in \mathbb{Z}_{>0}$, we also consider its version $W_{\mathrm{aff},k} = W_{\mathrm{fin}} \ltimes k\Lambda$; it acts naturally on $\Lambda_{\mathrm{aff},k} = \{k\} \times \Lambda \oplus \mathbb{Z}$ (trivially on $\mathbb{Z}$). Every $W_{\mathrm{aff},k}$-orbit on $\Lambda_{\mathrm{aff},k}$ contains a unique representative in $\Lambda_{\mathrm{aff},k}^{+}$. It follows that if we denote by Γ_k the group of roots of unity of order k, then there is a natural isomorphism $\Lambda_{\mathrm{aff},k}^{+}/\mathbb{Z} = W_{\mathrm{aff},k} \backslash \Lambda \xrightarrow{\sim} \mathrm{Hom}(\Gamma_k, G)/\mathrm{Ad}_G$.

4.2 The quest. We would like to have a double affine Grassmannian $\mathrm{Gr}_{G_{\mathrm{aff}}}$ and a geometric Satake equivalence between the category of integrable representations $\mathrm{Rep}(G_{\mathrm{aff}}^{\vee})$ and an appropriate category of perverse sheaves on $\mathrm{Gr}_{G_{\mathrm{aff}}}$. Note that the affine Satake isomorphism at the level of functions is established in Braverman and Kazhdan [2013] and Braverman, Kazhdan, and Patnaik [2016] (and in Gaussent and Rousseau [2014] for arbitrary Kac-Moody groups).

Such a quest was formulated by I. Grojnowski in his talk at ICM-2006 in Madrid. At approximately the same time, I. Frenkel suggested that the integrable representations of level k should be realized in cohomology of certain instanton moduli spaces on $\mathbb{A}^2/\Gamma_k$. Here Γ_k acts on $\mathbb{A}^2$ in a hyperbolic way: $\zeta(x, y) = (\zeta x, \zeta^{-1} y)$.

Note that the set of dominant coweights Λ^{+} is well ordered, which reflects the fact that the affine Grassmannian Gr_G is an ind-projective scheme. However, the set of affine dominant coweights $\Lambda_{\mathrm{aff}}^{+}$ is not well ordered: it does not have a minimal element. In fact, it has an automorphism group $\mathbb{Z}$ acting by the energy shifts: $(k, \bar{\lambda}, n) \mapsto (k, \bar{\lambda}, n + n')$ (we add a multiple of the minimal imaginary coroot δ). This indicates that the sought for double affine Grassmannian $\mathrm{Gr}_{G_{\mathrm{aff}}}$ is an object of semiinfinite nature.

At the moment, the only technical possibility of dealing with semiinfinite spaces is via transversal slices to strata. Following I. Frenkel's suggestion, in the series Braverman and

Finkelberg [2010, 2012, 2013] we developed a partial affine analogue of slices of Section 3 defined in terms of Uhlenbeck spaces $\mathfrak{U}_G(\mathbb{A}^2/\Gamma_k)$.

4.3 Dominant slices via Uhlenbeck spaces. The Uhlenbeck space $\mathfrak{U}_G^d(\mathbb{A}^2)$ is a partial closure of the moduli space $\mathrm{Bun}_G^d(\mathbb{A}^2)$ of G-bundles of second Chern class d on the projective plane $\mathbb{P}^2$ trivialized at the infinite line $\mathbb{P}^1_\infty \subset \mathbb{P}^2$, see Braverman, Finkelberg, and Gaitsgory [2006]. It is known that $\mathrm{Bun}_G^d(\mathbb{A}^2)$ is smooth quasiaffine, and $\mathfrak{U}_G^d(\mathbb{A}^2)$ is a connected affine variety of dimension $2dh_G^\vee$ (where $h_G^\vee$ is the dual Coxeter number of G). Conjecturally, $\mathfrak{U}_G^d(\mathbb{A}^2)$ is normal; in this case $\mathfrak{U}_G^d(\mathbb{A}^2)$ is the affinization of $\mathrm{Bun}_G^d(\mathbb{A}^2)$.

The group $G \times \mathrm{GL}(2)$ acts naturally on $\mathfrak{U}_G^d(\mathbb{A}^2)$: the first factor via the change of trivialization at $\mathbb{P}^1_\infty$, and the second factor via its action on $(\mathbb{P}^2, \mathbb{P}^1_\infty)$. The group Γ_k is embedded into GL(2). Given $\mu = (k, \bar{\mu}, m) \in \Lambda^+_{\mathrm{aff},k}$ we choose its lift to a homomorphism from Γ_k to G; thus Γ_k embeds diagonally into $G \times \mathrm{GL}(2)$ and acts on $\mathrm{Bun}_G^d(\mathbb{A}^2)$. The fixed point subvariety $\mathrm{Bun}_G^d(\mathbb{A}^2)^{\Gamma_k}$ consists of Γ_k-equivariant bundles and is denoted $\mathrm{Bun}_{G,\mu}^d(\mathbb{A}^2/\Gamma_k)$; another choice of lift above leads to an isomorphic subvariety. Since $0 \in \mathbb{A}^2$ is a Γ_k-fixed point, for any Γ_k-equivariant G-bundle $\mathcal{P} \in \mathrm{Bun}_{G,\mu}^d(\mathbb{A}^2/\Gamma_k)$ the group Γ_k acts on the fiber $\mathcal{P}_0$. This action defines an element of $\mathrm{Hom}(\Gamma_k, G)/\mathrm{Ad}_G$ to be denoted $[\mathcal{P}_0]$.

Now given $\lambda = (k, \bar{\lambda}, l) \in \Lambda^+_{\mathrm{aff},k}$ we define $\mathrm{Bun}_{G,\mu}^\lambda(\mathbb{A}^2/\Gamma_k)$ as the subvariety of $\mathrm{Bun}_{G,\mu}^d(\mathbb{A}^2/\Gamma_k)$ formed by all $\mathcal{P}$ such that the class $[\mathcal{P}_0] \in \mathrm{Hom}(\Gamma_k, G)/\mathrm{Ad}_G$ is the image of λ, and $d = k(l-m) + \frac{(\bar{\lambda},\bar{\lambda})-(\bar{\mu},\bar{\mu})}{2}$. It is a union of connected components of $\mathrm{Bun}_{G,\mu}^d(\mathbb{A}^2/\Gamma_k)$. Conjecturally, $\mathrm{Bun}_{G,\mu}^\lambda(\mathbb{A}^2/\Gamma_k)$ is connected. This conjecture is proved if $G = \mathrm{SL}(N)$, or $k = 1$, or k is big enough for arbitrary G and fixed $\bar{\lambda}, \bar{\mu}$.

Finally, we define the dominant slice $\overline{\mathcal{W}}{}_\mu^\lambda$ as the closure $\mathfrak{U}_{G,\mu}^\lambda(\mathbb{A}^2/\Gamma_k)$ of $\mathrm{Bun}_{G,\mu}^\lambda(\mathbb{A}^2/\Gamma_k)$ in the Uhlenbeck space $\mathfrak{U}_G^d(\mathbb{A}^2)$.

4.4 (Hyperbolic) stalks. The Cartan torus $T_{\mathrm{aff}} = \mathbb{C}^\times \times T \times \mathbb{C}^\times$ maps into $G \times \mathrm{GL}(2)$. Here the first copy of $\mathbb{C}^\times$ goes to the diagonal torus of $\mathrm{SL}(2) \subset \mathrm{GL}(2)$, while the second copy of $\mathbb{C}^\times$ goes to the center of GL(2). So T_{aff} acts on $\overline{\mathcal{W}}{}_\mu^\lambda$, and we denote by $\mu \in \overline{\mathcal{W}}{}_\mu^\lambda$ the only fixed point. The corresponding repellent R_μ^λ is the closed affine subvariety formed by all the points that flow into μ under the action of $2\rho(t)$, as t goes to ∞. The corresponding hyperbolic stalk $\Phi_\mu^\lambda \mathrm{IC}(\overline{\mathcal{W}}{}_\mu^\lambda)$ is conjecturally isomorphic to the weight space V_μ^λ of the integrable $G_{\mathrm{aff}}^\vee$-module V^λ with highest weight λ. In type A this conjecture follows from the identification of $\overline{\mathcal{W}}{}_\mu^\lambda$ with a Nakajima cyclic quiver variety and I. Frenkel's level-rank duality between the weight multiplicities and the tensor product multiplicities Frenkel [1982], Nakajima [2002, 2009], and Braverman and Finkelberg

[2010]. In type ADE at level 1 this conjecture follows from Braverman, Finkelberg, and Nakajima [2016b]. Also, as the notation suggests, the hyperbolic stalk $\Phi^\lambda_\mu \mathrm{IC}(\overline{\mathcal{W}}^\lambda_\mu)$ is isomorphic to the vanishing cycles of $\mathrm{IC}(\overline{\mathcal{W}}^\lambda_\mu)$ at μ with respect to a general function vanishing at μ Finkelberg and Kubrak [2015]. The usual stalk of $\mathrm{IC}(\overline{\mathcal{W}}^\lambda_\mu)$ at μ is conjecturally isomorphic to the associated graded of V^λ_μ with respect to the the affine Brylinski-Kostant filtration Slofstra [2012]. At level 1, this conjecture follows from the computation of the IC-stalks of Uhlenbeck spaces in Braverman, Finkelberg, and Gaitsgory [2006].

The affine analogs of generalized slices of Sections 3.2, 3.3, 3.4 were constructed in type A in Braverman and Finkelberg [2012, 2013] in terms of Nakajima cyclic quiver varieties mentioned above. For arbitrary G, the desired generalized slices are expected to be the Uhlenbeck partial compactifications of the moduli spaces of Γ_k-equivariant G_c-instantons (where $G_c \subset G$ is a maximal compact subgroup) on multi Taub-NUT spaces (for a physical explanation via a supersymmetric conformal field theory in 6 dimensions, see Witten [2010]). Unfortunately, we are still lacking a modular definition of the Uhlenbeck compactification Baranovsky [2015], and the existing *ad hoc* constructions are not flexible enough. Another approach via the Coulomb branches of framed affine quiver gauge theories following Nakajima [2016a] and Braverman, Finkelberg, and Nakajima [2016c,a, 2017] is described in the remaining sections. For a beautiful short introduction to the Coulomb branches, the reader may consult Nakajima [2016b, 2015].

5 Coulomb branches of $3d$ $\mathcal{N} = 4$ quiver gauge theories

5.1 General setup. Let $\mathbf{N}$ be a finite dimensional representation of a complex connected reductive group G (having nothing to do with G of previous sections). We consider the moduli space $\mathcal{R}_{\mathsf{G},\mathbf{N}}$ of triples $(\mathcal{P}, \sigma, s)$ where $\mathcal{P}$ is a G-bundle on the formal disc $D = \operatorname{Spec} \mathcal{O}$; σ is a trivialization of $\mathcal{P}$ on the punctured formal disc $D^* = \operatorname{Spec} \mathcal{K}$; and s is a section of the associated vector bundle $\mathcal{P}_{\mathrm{triv}} \overset{\mathsf{G}}{\times} \mathbf{N}$ on D^* such that s extends to a regular section of $\mathcal{P}_{\mathrm{triv}} \overset{\mathsf{G}}{\times} \mathbf{N}$ on D, and $\sigma(s)$ extends to a regular section of $\mathcal{P} \overset{\mathsf{G}}{\times} \mathbf{N}$ on D. In other words, s extends to a regular section of the vector bundle associated to the G-bundle glued from $\mathcal{P}$ and $\mathcal{P}_{\mathrm{triv}}$ on the non-separated formal scheme glued from 2 copies of D along D^* (*raviolo*). The group $\mathsf{G}_\mathcal{O}$ acts on $\mathcal{R}_{\mathsf{G},\mathbf{N}}$ by changing the trivialization σ, and we have an evident $\mathsf{G}_\mathcal{O}$-equivariant projection $\mathcal{R}_{\mathsf{G},\mathbf{N}} \to \mathrm{Gr}_\mathsf{G}$ forgetting s. The fibers of this projection are profinite dimensional vector spaces: the fiber over the base point is $\mathbf{N} \otimes \mathcal{O}$, and all the other fibers are subspaces in $\mathbf{N} \otimes \mathcal{O}$ of finite codimension. One may say that $\mathcal{R}_{\mathsf{G},\mathbf{N}}$ is a $\mathsf{G}_\mathcal{O}$-equvariant "constructible profinite dimensional vector bundle" over Gr_G. The $\mathsf{G}_\mathcal{O}$-equivariant Borel-Moore homology $H^{\mathsf{G}_\mathcal{O}}_\bullet(\mathcal{R}_{\mathsf{G},\mathbf{N}})$ is well-defined, and forms an associative algebra with respect to a convolution operation. This algebra is commutative, finitely generated and integral, and its spectrum $\mathcal{M}_C(\mathsf{G}, \mathbf{N}) = \operatorname{Spec} H_\bullet^{\mathsf{G}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$ is an irreducible normal affine variety of dimension $2\,\mathrm{rk}(\mathsf{G})$, the *Coulomb branch*. It is supposed to be a (singular) hyper-Kähler manifold Seiberg and Witten [1997].

Let $\mathsf{T} \subset \mathsf{G}$ be a Cartan torus with Lie algebra $\mathfrak{t} \subset \mathfrak{g}$. Let $\mathsf{W} = N_\mathsf{G}(\mathsf{T})/\mathsf{T}$ be the corresponding Weyl group. Then the equivariant cohomology $H^\bullet_{\mathsf{G}_\mathcal{O}}(\mathrm{pt}) = \mathbb{C}[\mathfrak{t}/\mathsf{W}]$ forms a subalgebra of $H_\bullet^{\mathsf{G}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$ (a *Cartan subalgebra*), so we have a projection $\Pi\colon \mathcal{M}_C(\mathsf{G}, \mathbf{N}) \to \mathfrak{t}/\mathsf{W}$.

Finally, the algebra $H_\bullet^{\mathsf{G}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$ comes equipped with quantization: a $\mathbb{C}[\hbar]$-deformation $\mathbb{C}_\hbar[\mathcal{M}_C(\mathsf{G}, \mathbf{N})] = H_\bullet^{\mathbb{C}^\times \ltimes \mathsf{G}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$ where $\mathbb{C}^\times$ acts by loop rotations, and $\mathbb{C}[\hbar] = H^\bullet_{\mathbb{C}^\times}(\mathrm{pt})$. It gives rise to a Poisson bracket on $\mathbb{C}[\mathcal{M}_C(\mathsf{G}, \mathbf{N})]$ with an open symplectic leaf, so that Π becomes an integrable system: $\mathbb{C}[\mathfrak{t}/\mathsf{W}] \subset \mathbb{C}[\mathcal{M}_C(\mathsf{G}, \mathbf{N})]$ is a Poisson-commutative polynomial subalgebra with $\mathrm{rk}(\mathsf{G})$ generators.

5.2 Flavor symmetry. Suppose we have an extension $1 \to \mathsf{G} \to \tilde{\mathsf{G}} \to \mathsf{G}_F \to 1$ where G_F is a connected reductive group (a *flavor group*), and the action of G on $\mathbf{N}$ is extended to an action of $\tilde{\mathsf{G}}$. Then the action of $\mathsf{G}_\mathcal{O}$ on $\mathcal{R}_{G,\mathbf{N}}$ extends to an action of $\tilde{\mathsf{G}}_\mathcal{O}$, and the convolution product defines a commutative algebra structure on the equivariant Borel-Moore homology $H_\bullet^{\tilde{\mathsf{G}}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$. We have the restriction homomorphism $H_\bullet^{\tilde{\mathsf{G}}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}}) \to H_\bullet^{\mathsf{G}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}}) = H_\bullet^{\tilde{\mathsf{G}}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}}) \otimes_{H^\bullet_{\mathsf{G}_F}(\mathrm{pt})} \mathbb{C}$. In other words, $\underline{\mathcal{M}}_C(\mathsf{G}, \mathbf{N}) := \operatorname{Spec} H_\bullet^{\tilde{\mathsf{G}}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$ is a deformation of $\mathcal{M}_C(\mathsf{G}, \mathbf{N})$ over $\operatorname{Spec} H^\bullet_{\mathsf{G}_F}(\mathrm{pt}) = \mathfrak{t}_F/\mathsf{W}_F$.

We will need the following version of this construction. Let $\mathsf{Z} \subset \mathsf{G}_F$ be a torus embedded into the flavor group. We denote by $\tilde{\mathsf{G}}^\mathsf{Z}$ the pullback extension $1 \to \mathsf{G} \to \tilde{\mathsf{G}}^\mathsf{Z} \to \mathsf{Z} \to 1$. We define $\underline{\mathcal{M}}_C^\mathsf{Z}(\mathsf{G}, \mathbf{N}) := \operatorname{Spec} H_\bullet^{\tilde{\mathsf{G}}^\mathsf{Z}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$: a deformation of $\mathcal{M}_C(\mathsf{G}, \mathbf{N})$ over $\mathfrak{z} := \operatorname{Spec} H^\bullet_\mathsf{Z}(\mathrm{pt})$.

Since $\mathcal{M}_C(\mathsf{G}, \mathbf{N})$ is supposed to be a hyper-Kähler manifold, its flavor deformation should come together with a (partial) resolution. To construct it, we consider the obvious projection $\tilde{\pi}\colon \mathcal{R}_{\tilde{\mathsf{G}},\mathbf{N}} \to \mathrm{Gr}_{\tilde{\mathsf{G}}} \to \mathrm{Gr}_{\mathsf{G}_F}$. Given a dominant coweight $\lambda_F \in \Lambda_F^+ \subset \mathrm{Gr}_{\mathsf{G}_F}$, we set $\mathcal{R}^{\lambda_F}_{\tilde{\mathsf{G}},\mathbf{N}} := \tilde{\pi}^{-1}(\lambda_F)$, and consider the equivariant Borel-Moore homology $H_\bullet^{\tilde{\mathsf{G}}^\mathsf{Z}_\mathcal{O}}(\mathcal{R}^{\lambda_F}_{\tilde{\mathsf{G}},\mathbf{N}})$. It carries a convolution module structure over $H_\bullet^{\tilde{\mathsf{G}}^\mathsf{Z}_\mathcal{O}}(\mathcal{R}_{\mathsf{G},\mathbf{N}})$. We consider

$$\underline{\widetilde{\mathcal{M}}}_C^{\mathsf{Z},\lambda_F}(\mathsf{G}, \mathbf{N}) := \operatorname{Proj}\Big(\bigoplus_{n\in\mathbb{N}} H_\bullet^{\tilde{\mathsf{G}}^\mathsf{Z}_\mathcal{O}}(\mathcal{R}^{n\lambda_F}_{\tilde{\mathsf{G}},\mathbf{N}})\Big) \xrightarrow{\varpi} \underline{\mathcal{M}}_C^\mathsf{Z}(\mathsf{G}, \mathbf{N}).$$

We denote $\varpi^{-1}(\mathfrak{M}_C(\mathsf{G},\mathbf{N}))$ by $\widetilde{\mathfrak{M}}_C^{\lambda_F}(\mathsf{G},\mathbf{N})$. We have

$$\widetilde{\mathfrak{M}}_C^{\lambda_F}(\mathsf{G},\mathbf{N}) = \operatorname{Proj}(\bigoplus_{n\in\mathbb{N}} H_\bullet^{\mathsf{G}_\mathcal{O}}(\mathcal{R}_{\tilde{\mathsf{G}},\mathbf{N}}^{n\lambda_F})).$$

More generally, for a strictly convex (i.e. not containing nontrivial subgroups) cone $\mathsf{V}\subset\Lambda_F^+$, we consider the multi projective spectra

$$\widetilde{\underline{\mathfrak{M}}}_C^{\mathsf{Z},\mathsf{V}}(\mathsf{G},\mathbf{N}) := \operatorname{Proj}(\bigoplus_{\lambda_F\in\mathsf{V}} H_\bullet^{\tilde{\mathsf{G}}_\mathcal{O}^{\mathsf{Z}}}(\mathcal{R}_{\tilde{\mathsf{G}},\mathbf{N}}^{\lambda_F})) \xrightarrow{\varpi} \underline{\mathfrak{M}}_C^{\mathsf{Z}}(\mathsf{G},\mathbf{N})$$

and

$$\widetilde{\mathfrak{M}}_C^{\mathsf{V}}(\mathsf{G},\mathbf{N}) := \operatorname{Proj}(\bigoplus_{\lambda_F\in\mathsf{V}} H_\bullet^{\mathsf{G}_\mathcal{O}}(\mathcal{R}_{\tilde{\mathsf{G}},\mathbf{N}}^{\lambda_F})) \xrightarrow{\varpi} \mathfrak{M}_C(\mathsf{G},\mathbf{N}).$$

5.3 Quiver gauge theories. Let Q be a quiver with Q_0 the set of vertices, and Q_1 the set of arrows. An arrow $e\in Q_1$ goes from its tail $t(e)\in Q_0$ to its head $h(e)\in Q_0$. We choose a Q_0-graded vector spaces $V := \bigoplus_{j\in Q_0} V_j$ and $W := \bigoplus_{j\in Q_0} W_j$. We set $\mathsf{G} = \mathrm{GL}(V) := \prod_{j\in Q_0}\mathrm{GL}(V_j)$. We choose a second grading $W = \bigoplus_{s=1}^N W^{(s)}$ compatible with the Q_0-grading of W. We set G_F to be a Levi subgroup $\prod_{s=1}^N\prod_{j\in Q_0}\mathrm{GL}(W_j^{(s)})$ of $\mathrm{GL}(W)$, and $\tilde{\mathsf{G}} := \mathsf{G}\times\mathsf{G}_F$. Finally, we define a central subgroup $\mathsf{Z}\subset\mathsf{G}_F$ as follows: $\mathsf{Z} := \prod_{s=1}^N\Delta_{\mathbb{C}^\times}^{(s)} \subset \prod_{s=1}^N\prod_{j\in Q_0}\mathrm{GL}(W_j^{(s)})$, where $\mathbb{C}^\times\cong\Delta_{\mathbb{C}^\times}^{(s)}\subset\prod_{j\in Q_0}\mathrm{GL}(W_j^{(s)})$ is the diagonally embedded subgroup of scalar matrices. The reductive group $\tilde{\mathsf{G}}$ acts naturally on $\mathbf{N} := \bigoplus_{e\in Q_1}\mathrm{Hom}(V_{t(e)},V_{h(e)})\oplus\bigoplus_{j\in Q_0}\mathrm{Hom}(W_j,V_j)$.

The Higgs branch of the corresponding quiver gauge theory is the Nakajima quiver variety $\mathfrak{M}_H(\mathsf{G},\mathbf{N}) = \mathfrak{M}(V,W)$. We are interested in the Coulomb branch $\mathfrak{M}_C(\mathsf{G},\mathbf{N})$.

5.4 Back to slices in an affine Grassmannian. Let now G be an adjoint simple simply laced algebraic group. We choose an orientation Ω of its Dynkin graph (of type ADE), and denote by I its set of vertices. Given an I-graded vector space W we encode its dimension by a dominant coweight $\bar{\lambda} := \sum_{i\in I}\dim(W_i)\bar{\omega}_i \in \Lambda^+$ of G. Given an I-graded vector space V we encode its dimension by a positive coroot combination $\alpha := \sum_{i\in I}\dim(V_i)\alpha_i \in \Lambda_+$. We set $\bar{\mu} := \bar{\lambda}-\alpha\in\Lambda$. Given a direct sum decomposition $W = \bigoplus_{s=1}^N W^{(s)}$ compatible with the I-grading of W as in Section 5.3, we set $\bar{\lambda}_s := \sum_{i\in I}\dim(W_i^{(s)})\bar{\omega}_i\in\Lambda^+$, and finally, $\underline{\bar{\lambda}} := (\bar{\lambda}_1,\dots,\bar{\lambda}_N)$.

Recall the notations of Section 5.2. Since the flavor group G_F is a Levi subgroup of $\mathrm{GL}(W)$, its weight lattice is naturally identified with $\mathbb{Z}^{\dim W}$. More precisely, we choose a basis $w_1,\dots,w_{\dim W}$ of W such that any W_i, $i\in I$, and $W^{(s)}$, $1\le s\le N$, is spanned by a subset of the basis, and we assume the following monotonicity condition: if for $1 \leq a < b < c \leq \dim W$ we have $w_a, w_b \in W^{(s)}$ for certain s, then $w_b \in W^{(s)}$ as well. We define a strictly convex cone $\mathsf{V} = \{(n_1, \dots, n_{\dim W})\} \subset \Lambda_F^+ \subset \mathbb{Z}^{\dim W}$ by the following conditions: (a) if $w_k \in W^{(s)}$, $w_l \in W^{(t)}$, and $s < t$, then $n_k \geq n_l \geq 0$; (b) if $w_k, w_l \in W^{(s)}$, then $n_k = n_l$. The following isomorphisms are constructed in Braverman, Finkelberg, and Nakajima [2016a] (notations of Section 3):

$$\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu} \xrightarrow{\sim} \mathfrak{M}_C(\mathsf{G}, \mathbf{N}),\ \underline{\overline{\mathcal{W}}}^{\bar\lambda}_{\bar\mu} \xrightarrow{\sim} \underline{\mathfrak{M}}^{\mathsf{Z}}_C(\mathsf{G}, \mathbf{N}),$$

(we learned of their existence from V. Pestun). We also expect the following isomorphisms:

$$\underline{\widetilde{\mathcal{W}}}^{\bar\lambda}_{\bar\mu} \xrightarrow{\sim} \underline{\widetilde{\mathfrak{M}}}^{\mathsf{Z},\mathsf{V}}_C(\mathsf{G}, \mathbf{N}),\ \widetilde{\mathcal{W}}^{\bar\lambda}_{\bar\mu} \xrightarrow{\sim} \widetilde{\mathfrak{M}}^{\mathsf{V}}_C(\mathsf{G}, \mathbf{N}).$$

In case G is an adjoint simple non simply laced algebraic group, it can be obtained by folding from a simple simply laced group $\tilde G$ (i.e. as the fixed point set of an outer automorphism of $\tilde G$). The corresponding automorphism of the Dynkin quiver of $\tilde G$ acts on the above Coulomb branches, and the slices for G can be realized as the fixed point sets of these Coulomb branches.

5.5 Back to slices in a double affine Grassmannian. We choose an orientation of an *affine* Dynkin graph of type $A^{(1)}, D^{(1)}, E^{(1)}$ with the set of vertices $\tilde I = I \sqcup \{i_0\}$, and repeat the construction of Section 5.4 for an affine dominant coweight $\lambda = \sum_{i \in \tilde I} \dim(W_i)\omega_i = (k, \bar\lambda, 0) \in \Lambda^+_{\mathrm{aff}}$, a positive coroot combination $\alpha = \sum_{i \in \tilde I} \dim(V_i)\alpha_i \in \Lambda_{\mathrm{aff},+}$, and $\mu := \lambda - \alpha = (k, \bar\mu, n) \in \Lambda_{\mathrm{aff}}$.

We define the slices in $\mathrm{Gr}_{G_{\mathrm{aff}}}$ (where G is the corresponding adjoint simple simply laced algebraic group) as

$$\overline{\mathcal{W}}^{\lambda}_{\mu} := \mathfrak{M}_C(\mathsf{G}, \mathbf{N}),\ \underline{\overline{\mathcal{W}}}^{\lambda}_{\mu} := \underline{\mathfrak{M}}^{\mathsf{Z}}_C(\mathsf{G}, \mathbf{N}),\ \underline{\widetilde{\mathcal{W}}}^{\lambda}_{\mu} := \underline{\widetilde{\mathfrak{M}}}^{\mathsf{Z},\mathsf{V}}_C(\mathsf{G}, \mathbf{N}),\ \widetilde{\mathcal{W}}^{\lambda}_{\mu} := \widetilde{\mathfrak{M}}^{\mathsf{V}}_C(\mathsf{G}, \mathbf{N}).$$

If μ is dominant, the slices $\overline{\mathcal{W}}^{\lambda}_{\mu}$ conjecturally coincide with the ones of Section 4.3. In type A this conjecture follows from the computation Nakajima and Takayama [2017] of Coulomb branches of the cyclic quiver gauge theories and their identification with the Nakajima cyclic quiver varieties.

Note that $\pi_0(\mathcal{R}_{\mathsf{G},\mathbf{N}}) = \pi_0(\mathrm{Gr}_{\mathrm{GL}(V)}) = \pi_1(\mathrm{GL}(V)) = \mathbb{Z}^{\tilde I}$, so that $H^{\mathsf{G}_\mathcal{O}}_\bullet(\mathcal{R}_{\mathsf{G},\mathbf{N}}) = \mathbb{C}[\mathfrak{M}_C(\mathsf{G}, \mathbf{N})] = \mathbb{C}[\overline{\mathcal{W}}^{\lambda}_{\mu}]$ is $\mathbb{Z}^{\tilde I}$-graded. We identify $\mathbb{Z}^{\tilde I}$ with the root lattice of $T_{\mathrm{aff}} \subset G_{\mathrm{aff}}\colon \mathbb{Z}^{\tilde I} = \mathbb{Z}\langle\alpha_i^\vee\rangle_{i \in \tilde I}$. Then the $\mathbb{Z}^{\tilde I}$-grading on $\mathbb{C}[\overline{\mathcal{W}}^{\lambda}_{\mu}]$ corresponds to a T_{aff}-action on $\overline{\mathcal{W}}^{\lambda}_{\mu}$. Composing with the cocharacter $2\rho\colon \mathbb{C}^\times \to T_{\mathrm{aff}}$, we obtain an action of $\mathbb{C}^\times$ on $\overline{\mathcal{W}}^{\lambda}_{\mu}$. Conjecturally, the fixed point set $(\overline{\mathcal{W}}^{\lambda}_{\mu})^{\mathbb{C}^\times}$ is nonempty iff the $V^{\lambda}_{\mu} \neq 0$, and in this case the fixed point set consists of a single point denoted by μ. We consider the corresponding repellent $R^\lambda_\mu \subset \overline{\mathcal{W}}^\lambda_\mu$ and the hyperbolic stalk $\Phi^\lambda_\mu \mathrm{IC}(\overline{\mathcal{W}}^\lambda_\mu)$.

Similarly to Section 5.4, in case G is an adjoint simple non simply laced group, the Dynkin diagram of its affinization can be obtained by folding of a Dynkin graph of type $A^{(1)}, D^{(1)}, E^{(1)}$, and the above slices for G are defined as the fixed point sets of the corresponding slices for the unfolding of G. The repellents and the hyperbolic stalks are thus defined for arbitrary simple G too, and we expect the conclusions of Sections 3.2, 3.3, 3.4 to hold in the affine case as well.

5.6 Warning. In order to formulate the statements about multiplicities for fusion and convolution as in 3.3 and 3.4, we must have closed embeddings of slices $\overline{\mathcal{W}}^{\lambda'}_\mu \hookrightarrow \overline{\mathcal{W}}^\lambda_\mu$ for $\lambda' \le \lambda \in \Lambda^+_{\mathrm{aff}}$. Certainly we do have the natural closed embeddings of generalized slices in $\mathrm{Gr}_G : \overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu} \hookrightarrow \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$, $\bar\lambda' \le \bar\lambda \in \Lambda^+$, but these embeddings have no manifest interpretation in terms of Coulomb branches (see Section 6.4 below for a partial advance, though). For a slice in Gr_G, the collection of closures of symplectic leaves in $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ coincides with the collection of smaller slices $\overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu} \subset \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$, $\bar\mu \le \bar\lambda' \le \bar\lambda$, $\bar\lambda' \in \Lambda^+$. However, in the affine case, in general there are *more* symplectic leaves in $\overline{\mathcal{W}}^\lambda_\mu$ than the cardinality of $\{\lambda' \in \Lambda^+_{\mathrm{aff}} : \mu \le \lambda' \le \lambda\}$. For example, if $k = 1$, and $\mu = 0$, so that $\overline{\mathcal{W}}^\lambda_\mu \simeq \mathcal{U}^d_G(\mathbb{A}^2)$, the symplectic leaves are numbered by the partitions of size $\le d$: they are all of the form $\mathcal{S} \times \mathrm{Bun}^{d'}_G(\mathbb{A}^2)$ where $0 \le d' \le d$, and $\mathcal{S}$ is a stratum of the diagonal stratification of $\mathrm{Sym}^{d-d'} \mathbb{A}^2$.

Thus we expect that the slice $\overline{\mathcal{W}}^{\lambda'}_\mu$ for $\lambda' \in \Lambda^+_{\mathrm{aff}}$, $\mu \le \lambda' \le \lambda$, is isomorphic to the closure of a symplectic leaf in $\overline{\mathcal{W}}^\lambda_\mu$. We also do expect the multiplicity of $\mathrm{IC}(\overline{\mathcal{W}}^{\lambda'}_\mu)$ in $\Psi\mathrm{IC}(\underline{\overline{\mathcal{W}}}^{\underline{\lambda}}_\mu) = \varpi_*\mathrm{IC}(\widetilde{\mathcal{W}}^{\underline{\lambda}}_\mu)$ to be $M^{\underline{\lambda}}_{\lambda'} = \mathrm{Hom}_{G^\vee_{\mathrm{aff}}}(V^{\lambda'}, V^{\lambda_1} \otimes \ldots \otimes V^{\lambda_N})$ for any $\lambda' \in \Lambda^+_{\mathrm{aff}}$ such that $\mu \le \lambda' \le \lambda$. However, it is possible that the IC sheaves of *other* symplectic leaves' closures also enter $\Psi\mathrm{IC}(\underline{\overline{\mathcal{W}}}^{\underline{\lambda}}_\mu) = \varpi_*\mathrm{IC}(\widetilde{\mathcal{W}}^{\underline{\lambda}}_\mu)$ with nonzero multiplicities. We should understand the representation-theoretic meaning of these extra multiplicities, cf. Nakajima [2009, Theorem 5.15 and Remark 5.17(3)] for G of type A.

Also, the closed embeddings of slices (for Levi subgroups of G_{aff}) seem an indispensable tool for constructing a $\mathfrak{g}^\vee_{\mathrm{aff}}$-action on $\bigoplus_\mu \Phi^\lambda_\mu \mathrm{IC}(\overline{\mathcal{W}}^\lambda_\mu)$ or a structure of $\mathfrak{g}^\vee_{\mathrm{aff}}$-crystal on $\bigsqcup_\mu \mathrm{Irr}(R^\lambda_\mu)$ (via reduction to Levi subgroups), cf. Krylov [2017].

5.7 Further problems. Note that the construction of Section 5.5 uses no specific properties of the affine Dynkin graphs, and works in the generality of arbitrary graph Q without edge loops and the corresponding Kac-Moody Lie algebra $\mathfrak{g}_Q$. We still expect the conclusions of Sections 3.2, 3.3, 3.4 to hold in this generality, see Braverman, Finkelberg, and Nakajima [2016a, 3(x)].

The only specific feature of the affine case is as follows. Recall that the category $\mathrm{Rep}_k(G^\vee_{\mathrm{aff}})$ of integrable $G^\vee_{\mathrm{aff}}$-modules at level $k \in \mathbb{Z}_{>0}$ is equipped with a braided balanced tensor *fusion* structure Moore and Seiberg [1989] and Bakalov and Kirillov [2001]. Unfortunately, I have no clue how this structure is reflected in the geometry of $\mathrm{Gr}_{G_{\mathrm{aff}}}$. I believe this is one of the most pressing problems about $\mathrm{Gr}_{G_{\mathrm{aff}}}$.

6 Applications

6.1 Hikita conjecture. We already mentioned in Section 5.5 that in case $V^\lambda_\mu \neq 0$ we expect the fixed point set $(\overline{\mathcal{W}}^\lambda_\mu)^{T_{\mathrm{aff}}}$ to consist of a single point μ. This point is the support of a nilpotent scheme $\boldsymbol{\mu}$ defined as follows: we choose a T_{aff}-equivariant embedding $\overline{\mathcal{W}}^\lambda_\mu \hookrightarrow \mathbb{A}^N$ into a representation of T_{aff}, and define $\boldsymbol{\mu}$ as the scheme-theoretic intersection of $\overline{\mathcal{W}}^\lambda_\mu$ with the zero weight subspace $\mathbb{A}^N_0$ inside $\mathbb{A}^N$. The resulting subscheme $\boldsymbol{\mu} \subset \overline{\mathcal{W}}^\lambda_\mu$ is independent of the choice of a T_{aff}-equivariant embedding $\overline{\mathcal{W}}^\lambda_\mu \hookrightarrow \mathbb{A}^N$. According to the Hikita conjecture Hikita [2017], the ring $\mathbb{C}[\boldsymbol{\mu}]$ is expected to be isomorphic to the cohomology ring $H^\bullet(\mathfrak{M}(V,W))$ of the corresponding Nakajima affine quiver variety, see Section 5.3. This is an instance of *symplectic duality* ($3d$ mirror symmetry) between Coulomb and Higgs branches. The Hikita conjecture for the slices $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ in Gr_G and the corresponding finite type Nakajima quiver varieties is proved in Kamnitzer, Tingley, Webster, Weekes, and Yacobi [2015] for types A, D (and conditionally for types E).

6.2 Monopole formula. We return to the setup of Section 5.1. Recall that $\mathcal{R}_{\mathbb{G},\mathbf{N}}$ is a union of (profinite dimensional) vector bundles over $\mathbb{G}_{\mathcal{O}}$-orbits in $\mathrm{Gr}_{\mathbb{G}}$. The corresponding Cousin spectral sequence converging to $H^{\mathbb{G}_{\mathcal{O}}}_\bullet(\mathcal{R}_{\mathbb{G},\mathbf{N}})$ degenerates and allows to compute the equivariant Poincaré polynomial (or rather Hilbert series)

$$P^{\mathbb{G}_{\mathcal{O}}}_t(\mathcal{R}_{\mathbb{G},\mathbf{N}}) = \sum_{\theta \in \Lambda^+_{\mathbb{G}}} t^{d_\theta - 2\langle \rho_{\mathbb{G}}, \theta\rangle} P_{\mathbb{G}}(t;\theta). \tag{1}$$

Here $\deg(t) = 2$, $P_{\mathbb{G}}(t;\theta) = \prod (1-t^{d_i})^{-1}$ is the Hilbert series of the equivariant cohomology $H^\bullet_{\mathrm{Stab}_{\mathbb{G}}(\theta)}(\mathrm{pt})$ (d_i are the degrees of generators of the ring of $\mathrm{Stab}_{\mathbb{G}}(\theta)$-invariant functions on its Lie algebra), and $d_\theta = \sum_{\chi \in \Lambda^\vee_{\mathbb{G}}} \max(-\langle \chi, \theta\rangle, 0) \dim \mathbf{N}_\chi$. This is a slight variation of the *monopole formula* of Cremonesi, Hanany, and Zaffaroni [2014]. Note that the series (1) may well diverge (even as a formal Laurent series: the space of homology of given degree may be infinite-dimensional), e.g. this is always the case for unframed quiver gauge theories. To ensure its convergence (as a formal Taylor series with the constant term 1) one has to impose the so called 'good' or 'ugly' assumption on the theory. In this case the resulting $\mathbb{N}$-grading on $H_\bullet^{\mathbb{G}_\mathcal{O}}(\mathcal{R}_{\mathbb{G},\mathbf{N}})$ gives rise to a $\mathbb{C}^\times$-action on $\mathfrak{M}_C(\mathbb{G},\mathbf{N})$, making it a conical variety with a single (attracting) fixed point.

Now recall the setup of Sections 5.3, 5.4; in particular, the isomorphism $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu} \xrightarrow{\sim} \mathfrak{M}_C(\mathbb{G},\mathbf{N})$. In case $\bar\mu$ is dominant, the slice $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu} \subset \mathrm{Gr}_G$ is conical with respect to the loop rotation $\mathbb{C}^\times$-action. However, this action is *not* the one of the previous paragraph. They differ by a hamiltonian $\mathbb{C}^\times$-action (preserving the Poisson structure). The Hilbert series of $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ graded by the loop rotation $\mathbb{C}^\times$-action is given by

$$P_t(\mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}]) = \sum_{\theta\in\Lambda^+_{\mathbb{G}}} t^{d_\theta - 2\langle\rho_{\mathbb{G}},\theta\rangle - \frac{1}{2}\bar\theta^\dagger\cdot\det\mathbf{N}_{\mathrm{hor}} + \frac{1}{2}\bar\theta^\dagger\cdot C\cdot\alpha} P_{\mathbb{G}}(t;\theta). \tag{2}$$

Here $\deg(t) = 1$; $\alpha = \bar\lambda - \bar\mu \in \Lambda_+ = \mathbb{N}^I$; $\bar\theta$ is the class of $\theta \in \Lambda_{\mathbb{G}} = \Lambda_{\mathrm{GL}(V)}$ in $\pi_0\mathrm{Gr}_{\mathrm{GL}(V)} = \mathbb{Z}^I$; $\bar\theta^\dagger$ is the transposed row-vector; C is the $I\times I$ Cartan matrix of G; and $\mathbf{N}_{\mathrm{hor}} = \bigoplus_{i\to j\in\Omega}\mathrm{Hom}(V_i,V_j)$ is the "horizontal" summand of $\mathrm{GL}(V)$-module $\mathbf{N}$, so that $\det\mathbf{N}_{\mathrm{hor}}$ is a character of $\mathrm{GL}(V)$, i.e. an element of $\mathbb{Z}^I$.

Finally, we consider a double affine Grassmannian slice $\overline{\mathcal{W}}^\lambda_\mu$ with dominant μ as in Section 4.3. The analogue of the loop rotation action of the previous paragraph is the action of the second copy of $\mathbb{C}^\times$ (the center of $\mathrm{GL}(2)$) in Section 4.4. We expect that the Hilbert series of $\overline{\mathcal{W}}^\lambda_\mu$ graded by this $\mathbb{C}^\times$-action is given by the evident affine analogue of the formula (2) (with the $\tilde I\times\tilde I$ Cartan matrix C_{aff} of G_{aff} replacing C). In particular, in case of level 1, this gives a formula for the Hilbert series of the coordinate ring $\mathbb{C}[\mathcal{U}^d_G(\mathbb{A}^2)]$ of the Uhlenbeck space proposed in Cremonesi, Mekareeya, Hanany, and Ferlito [2014]. Note that the latter formula works for arbitrary G, not necessarily simply laced one. In type A it follows from the results of Nakajima and Takayama [2017].

6.3 Zastava.

Let us consider the Coulomb branch $\mathfrak{M}_C(\mathbb{G},\mathbf{N})$ of an *unframed* quiver gauge theory for an ADE type quiver: $W_i = 0\ \forall i\in I$, so that $\mathbf{N} = \mathbf{N}_{\mathrm{hor}}$. An isomorphism $\mathfrak{M}_C(\mathbb{G},\mathbf{N}) \xrightarrow{\sim} \mathring{Z}^\alpha$ with the open zastava[1] (the moduli space of degree α based maps from the projective line $\mathbb{P}^1 \ni \infty$ to the flag variety $\mathcal{B} \ni B_-$ of G, where $\alpha = \sum_{i\in I}(\dim V_i)\alpha_i$), is constructed in Braverman, Finkelberg, and Nakajima [2016a] (we learned of its existence from V. Pestun). As the name suggests, the open zastava is a (dense smooth symplectic) open subvariety in the zastava space Z^α, a normal Cohen-Macaulay affine Poisson variety.

Note that there is another version of zastava $\mathbf{Z}^\alpha$ that is the solution of a moduli problem (G-bundles on $\mathbb{P}^1$ with a generalized B-structure and an extra U_--structure transversal at $\infty \in \mathbb{P}^1$) Braverman, Finkelberg, Gaitsgory, and Mirković [2002] given by a scheme cut out by the Plücker equations. This scheme is not reduced in general (the first example

[1] Zastava = *flags* in Croatian.

occurs in type A_4) E. Feigin and Makedonskyi [2017], and Z^α is the corresponding variety: $Z^\alpha := \mathbf{Z}^\alpha_{\mathrm{red}}$.

We already mentioned that the open zastava is a particular case of a generalized slice: $\mathring{Z}^\alpha = \overline{\mathcal{W}}^0_{w_0\alpha}$. The zastava space Z^α is the limit of slices in the following sense: for any $\bar\lambda \geq \bar\mu$ such that $w_0\bar\mu - w_0\bar\lambda = \alpha$, there is a loop rotation equivariant regular birational morphism $s^{\bar\lambda}_{\bar\mu}\colon \overline{\mathcal{W}}^{-w_0\bar\lambda}_{-w_0\bar\mu} \to Z^\alpha$, and for any $N \in \mathbb{N}$ and big enough dominant $\bar\mu$, the corresponding morphism of the coordinate rings graded by the loop rotations $(s^{\bar\lambda}_{\bar\mu})^*\colon \mathbb{C}[Z^\alpha] \to \mathbb{C}[\overline{\mathcal{W}}^{-w_0\bar\lambda}_{-w_0\bar\mu}]$ is an isomorphism in degrees $\leq N$ (both $\mathbb{C}[Z^\alpha]$ and $\mathbb{C}[\overline{\mathcal{W}}^{-w_0\bar\lambda}_{-w_0\bar\mu}]$ for dominant $\bar\mu$ are positively graded).

Now $\mathbb{C}[Z^\alpha]$ is obtained by the following version of the Coulomb branch construction. Given a vector space U we define the positive part of the affine Grassmannian $\mathrm{Gr}^+_{GL(U)} \subset \mathrm{Gr}_{GL(U)}$ as the moduli space of vector bundles $\mathcal{U}$ on the formal disc $D = \mathrm{Spec}(\mathcal{O})$ equipped with trivialization $\sigma\colon \mathcal{U}|_{D^*} \xrightarrow{\sim} U \otimes \mathcal{O}_{D^*}$ on the formal punctured disc $D^* = \mathrm{Spec}(\mathcal{K})$ such that σ extends through the puncture as an embedding $\sigma\colon \mathcal{U} \hookrightarrow U \otimes \mathcal{O}_D$. Since $\mathbb{G} = \mathrm{GL}(V) = \prod_{i\in I} \mathrm{GL}(V_i)$, we have $\mathrm{Gr}_{\mathrm{GL}(V)} = \prod_{i\in I} \mathrm{Gr}_{\mathrm{GL}(V_i)}$, and we define $\mathrm{Gr}^+_{\mathrm{GL}(V)} = \prod_{i\in I} \mathrm{Gr}^+_{\mathrm{GL}(V_i)}$. Finally, we define $\mathcal{R}^+_{\mathbb{G},\mathbf{N}}$ as the preimage of $\mathrm{Gr}^+_{\mathrm{GL}(V)} \subset \mathrm{Gr}_{\mathrm{GL}(V)}$ under $\mathcal{R}_{\mathbb{G},\mathbf{N}} \to \mathrm{Gr}_{\mathrm{GL}(V)}$. Then $H^{\mathbb{G}_\mathcal{O}}_\bullet(\mathcal{R}^+_{\mathbb{G},\mathbf{N}})$ forms a convolution subalgebra of $H^{\mathbb{G}_\mathcal{O}}_\bullet(\mathcal{R}_{\mathbb{G},\mathbf{N}})$, and an isomorphism $\mathcal{M}^+_C(\mathbb{G},\mathbf{N}) := \mathrm{Spec}\, H^{\mathbb{G}_\mathcal{O}}_\bullet(\mathcal{R}^+_{\mathbb{G},\mathbf{N}}) \xrightarrow{\sim} Z^\alpha$ is constructed in Braverman, Finkelberg, and Nakajima [2016a].

An analogue of the monopole formula (2) gives the character of the $T \times \mathbb{C}^\times$-module $\mathbb{C}[Z^\alpha]$:

$$\chi(\mathbb{C}[Z^\alpha]) = \sum_{\Lambda^{++}_{\mathbb{G}}} z^{\bar\theta} t^{d_\theta - 2\langle\rho_{\mathbb{G}},\theta\rangle - \frac{1}{2}\bar\theta^\dagger\cdot\det\mathbf{N} + \frac{1}{2}\bar\theta^\dagger\cdot C\cdot\alpha} P_{\mathbb{G}}(t;\theta). \tag{3}$$

Here $\Lambda^{++}_{\mathbb{G}}$ is the set of I-tuples of partitions; i-th partition having length at most $\dim V_i$ (recall that the cone of dominant coweights $\Lambda^+_{\mathbb{G}}$ is formed by the I-tuples of nonincreasing sequences $(\lambda^{(i)}_1 \geq \lambda^{(i)}_2 \geq \ldots \geq \lambda^{(i)}_{\dim V_i})$ of integers, and for $\Lambda^{++}_{\mathbb{G}} \subset \Lambda^+_{\mathbb{G}}$ we require these integers to be nonnegative). Also, z denotes the coordinates on the Cartan torus $T \subset G$ identified with $(\mathbb{C}^\times)^I$ via $z_i = \alpha^\vee_i$.

The character of the $T \times \mathbb{C}^\times$-module $\mathbb{C}[Z^\alpha]$ for G of type ADE was also computed in Braverman and Finkelberg [2014]. Namely, it is given by the *fermionic formula* of B. Feigin, E. Feigin, Jimbo, Miwa, and Mukhin [2009], and the generating function of these characters for all $\alpha \in \Lambda_+$ is an eigenfunction of the q-difference Toda integrable system. It would be interesting to find a combinatorial relation between the monopole and fermionic formulas.

In the affine case, the zastava space $Z^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ was introduced in Braverman, Finkelberg, and Gaitsgory [2006]. It is an irreducible affine algebraic variety containing a (dense smooth symplectic) open subvariety $\mathring{Z}^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$: the moduli space of degree α based maps from the projective line $\mathbb{P}^1$ to the Kashiwara flag scheme $\mathbf{Fl}_{\mathfrak{g}_{\mathrm{aff}}}$. Contrary to the finite case, the open subvariety $\mathring{Z}^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ is not affine, but only quasiaffine, and we denote by $\underline{\mathring{Z}}^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ its affine closure. We do not know if the open embedding $\mathring{Z}^\alpha_{\mathfrak{g}_{\mathrm{aff}}} \hookrightarrow Z^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ extends to an open embedding $\underline{\mathring{Z}}^\alpha_{\mathfrak{g}_{\mathrm{aff}}} \hookrightarrow Z^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$: it depends on the normality property of $Z^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ that is established only for $\mathfrak{g}$ of types A, C Braverman and Finkelberg [2014] and Finkelberg and Rybnikov [2014] at the moment (but is expected for all types). For an $A^{(1)}, D^{(1)}, E^{(1)}$ type quiver and an unframed quiver gauge theory with $\alpha = \sum_{i \in \tilde{I}} (\dim V_i)\alpha_i$, we have an isomorphism $\mathcal{M}_C(\mathsf{G}, \mathbf{N}) \xrightarrow{\sim} \underline{\mathring{Z}}^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ Braverman, Finkelberg, and Nakajima [2016a]. If $Z^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$ is normal, this isomorphism extends to $\mathcal{M}_C^+(\mathsf{G}, \mathbf{N}) \xrightarrow{\sim} Z^\alpha_{\mathfrak{g}_{\mathrm{aff}}}$, and the fermionic formula for the character $\chi(\mathbb{C}[Z^\alpha])$ holds true.

Finally, for an arbitrary quiver Q without edge loops we can consider an unframed quiver gauge theory, and a coroot $\alpha := \sum_{i \in Q_0} (\dim V_i)\alpha_i$ of the corresponding Kac-Moody Lie algebra $\mathfrak{g}_Q$. The moduli space $\mathring{Z}^\alpha_{\mathfrak{g}_Q}$ of based maps from $\mathbb{P}^1$ to the Kashiwara flag scheme $\mathbf{Fl}_{\mathfrak{g}_Q}$ was studied in Braverman, Finkelberg, and Gaitsgory [2006]. It is a smooth connected variety. We expect that it is quasiaffine, and its affine closure $\underline{\mathring{Z}}^\alpha_{\mathfrak{g}_Q}$ is isomorphic to the Coulomb branch $\mathcal{M}_C(\mathsf{G}, \mathbf{N})$. It would be interesting to find a modular interpretation of $\mathcal{M}_C^+(\mathsf{G}, \mathbf{N})$ and its stratification into symplectic leaves. In the affine case such an interpretation involves Uhlenbeck spaces $\mathfrak{U}^d_G(\mathbb{A}^2)$.

The Jordan quiver corresponds to the Heisenberg Lie algebra. The computations of Finkelberg, Ginzburg, Ionov, and Kuznetsov [2016] suggest that the Uhlenbeck compactification of the Calogero-Moser phase space plays the role of zastava for the Heisenberg Lie algebra.

6.4 Multiplication and quantization. The multiplication of slices in the affine Grassmannian $\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu} \times \overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu'} \to \overline{\mathcal{W}}^{\bar\lambda+\bar\lambda'}_{\bar\mu+\bar\mu'}$ was constructed in Braverman, Finkelberg, and Nakajima [2016a] via multiplication of scattering matrices for singular monopoles (we learned of its existence from T. Dimofte, D. Gaiotto and J. Kamnitzer). The corresponding comultiplication $\mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda+\bar\lambda'}_{\bar\mu+\bar\mu'}] \to \mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}] \otimes \mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu'}]$ can not be seen directly from the Coulomb branch construction of slices. However, its quantization $\mathbb{C}_\hbar[\overline{\mathcal{W}}^{\bar\lambda+\bar\lambda'}_{\bar\mu+\bar\mu'}] \to \mathbb{C}_\hbar[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}] \otimes \mathbb{C}_\hbar[\overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu'}]$ (recall from the end of Section 5.1 that $\mathbb{C}_\hbar[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}]$ is the loop rotation equivariant Borel-Moore homology of the corresponding variety of triples) already can be realized in terms of Coulomb branches. The reason for this is that the quantized Coulomb branch $\mathbb{C}_\hbar[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}]$ is likely to have a presentation by generators and relations of a *truncated shifted Yangian* $Y^{\bar\lambda}_{\bar\mu} \simeq \mathbb{C}_\hbar[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}]$ of Braverman, Finkelberg, and Nakajima [ibid., Appendix B]. Also, it seems likely that the comultiplication of Finkelberg, Kamnitzer, Pham, Rybnikov, and

Weekes [2018] descends to a homomorphism $\Delta\colon Y^{\bar\lambda+\bar\lambda'}_{\bar\mu+\bar\mu'} \to Y^{\bar\lambda}_{\bar\mu} \otimes Y^{\bar\lambda'}_{\bar\mu'}$. Finally, we expect that the desired comultiplication $\mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda+\bar\lambda'}_{\bar\mu+\bar\mu'}] \to \mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}] \otimes \mathbb{C}[\overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu'}]$ is obtained by setting $\hbar = 0$ in Δ.

Returning to the question of constructing the closed embeddings of slices $\overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu} \hookrightarrow \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ in terms of Coulomb branches (see the beginning of Section 5.6), we choose dominant coweights $\bar\nu, \bar\nu', \bar\mu'$ such that $\bar\nu' + \bar\nu = \bar\lambda'$, $\bar\mu' + \bar\nu = \bar\lambda'$, and set $\bar\mu'' := \bar\mu - \bar\mu'$. Then we have the multiplication morphism $\overline{\mathcal{W}}^{\bar\nu}_{\bar\mu''} \times \overline{\mathcal{W}}^{\bar\nu'}_{\bar\mu'} \to \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$, and we restrict it to $\overline{\mathcal{W}}^{\bar\nu}_{\bar\mu''} = \overline{\mathcal{W}}^{\bar\nu}_{\bar\mu''} \times \{\bar\mu'\} \to \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ where $\bar\mu' \in \overline{\mathcal{W}}^{\bar\nu'}_{\bar\mu'}$ is the fixed point. Then the desired closed subvariety $\overline{\mathcal{W}}^{\bar\lambda'}_{\bar\mu} \subset \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$ is nothing but the closure of the image of $\overline{\mathcal{W}}^{\bar\nu}_{\bar\mu''} \to \overline{\mathcal{W}}^{\bar\lambda}_{\bar\mu}$.

The similar constructions are supposed to work for the slices in $\mathrm{Gr}_{G_{\mathrm{aff}}}$. They are based on the comultiplication for *affine Yangians* constructed in Guay, Nakajima, and Wendlandt [2017].

6.5 Affinization of KLR algebras. We recall the setup of Section 5.3, and set $W = 0$ (no framing). We choose a sequence $\mathbf{j} = (j_1, \dots, j_N)$ of vertices such that any vertex $j \in Q_0$ enters $\dim V_j$ times; thus, $N = \dim V$. The set of all such sequences is denoted $\mathbf{J}(V)$. We choose a Q_0-graded flag $V = V^0 \supset V^1 \supset \dots \supset V^N = 0$ such that V^{n-1}/V^n is a 1-dimensional vector space supported at the vertex j_n for any $n = 1, \dots, N$. It gives rise to the following flag of Q_0-graded lattices in $V_\mathcal{K} = V \otimes \mathcal{K}$: $\dots \supset L^{-1} \supset L^0 \supset L^1 \supset \dots$, where $L^{r+N} = zL^r$ for any $r \in \mathbb{Z}$; $L^0 = V_\mathcal{O}$, and $L^n/L^N = V^n \subset V = L^0/L^N$ for any $n = 1, \dots, N$. Let $\mathsf{I}_\mathbf{j} \subset \mathsf{G}_\mathcal{O}$ be the stabilizer of the flag $L^\bullet$ (an Iwahori subgroup). Then the $\mathsf{I}_\mathbf{j}$-module $\mathbf{N}_\mathcal{O}$ contains a submodule $\mathbf{N}_\mathbf{j}$ formed by the $\mathcal{K}$-linear homomorphisms $b_e\colon V_{t(e),\mathcal{K}} \to V_{h(e),\mathcal{K}}$ such that for any $e \in Q_1$ and $r \in \mathbb{Z}$, b_e takes $L^r_{t(e)}$ to $L^{r+1}_{h(e)}$.

We consider the following version of the variety of triples: $\mathcal{R}_{\mathbf{j},\mathbf{j}} := \{[g, s] \in \mathsf{G}_\mathcal{K} \overset{\mathsf{I}_\mathbf{j}}{\times} \mathbf{N}_\mathbf{j} : gs \in \mathbf{N}_\mathbf{j}\}$, cf. Braverman, Etingof, and Finkelberg [2016] and Webster [2016, Section 4]. Then the equivariant Borel-Moore homology $\mathcal{H}_{\mathbf{j},\mathbf{j}} := H^{\mathbb{C}^\times \ltimes \mathsf{I}_\mathbf{j}}_\bullet(\mathcal{R}_{\mathbf{j},\mathbf{j}})$ forms an associative algebra with respect to a convolution operation. Moreover, if we take another sequence $\mathbf{j}' \in \mathbf{J}(V)$ and consider $\mathcal{R}_{\mathbf{j}',\mathbf{j}} := \{[g, s] \in \mathsf{G}_\mathcal{K} \overset{\mathsf{I}_\mathbf{j}}{\times} \mathbf{N}_\mathbf{j} : gs \in \mathbf{N}_{\mathbf{j}'}\}$, then $\mathcal{H}_{\mathbf{j}',\mathbf{j}} := H^{\mathbb{C}^\times \ltimes \mathsf{I}_{\mathbf{j}'}}_\bullet(\mathcal{R}_{\mathbf{j},\mathbf{j}'})$ forms a $\mathcal{H}_{\mathbf{j}',\mathbf{j}'} - \mathcal{H}_{\mathbf{j},\mathbf{j}}$-bimodule with respect to convolution, and we have convolutions $\mathcal{H}_{\mathbf{j}'',\mathbf{j}'} \otimes \mathcal{H}_{\mathbf{j}',\mathbf{j}} \to \mathcal{H}_{\mathbf{j}'',\mathbf{j}}$. In other words, $\mathcal{H}_V := \bigoplus_{\mathbf{j},\mathbf{j}' \in \mathbf{J}(V)} \mathcal{H}_{\mathbf{j}',\mathbf{j}}$ forms a convolution algebra.

Furthermore, given $\mathbf{j}_1 \in \mathbf{J}(V)$, $\mathbf{j}_2 \in \mathbf{J}(V')$, the concatenated sequence $\mathbf{j}_1\mathbf{j}_2$ lies in $\mathbf{J}(V \oplus V')$, and one can define the morphisms $\mathcal{H}_{\mathbf{j}'_1,\mathbf{j}_1} \otimes \mathcal{H}_{\mathbf{j}'_2,\mathbf{j}_2} \to \mathcal{H}_{\mathbf{j}'_1\mathbf{j}'_2,\mathbf{j}_1\mathbf{j}_2}$ summing up to a homomorphism $\mathcal{H}_V \otimes \mathcal{H}_{V'} \to \mathcal{H}_{V \oplus V'}$.

Similarly to the classical theory of Khovanov-Lauda-Rouquier algebras (see a beautiful survey Rouquier [2012] and references therein), we expect that in case Q has no loop edges, the categories of finitely generated graded projective $\mathcal{H}_V$-modules provide a categorification of the positive part U_Q^{++} of the quantum toroidal algebra U_Q (where U_Q^{++} is defined as the subalgebra generated by the positive modes of the positive generators $e_{j,r} : j \in Q_0,\ r \geq 0$).

References

Bojko Bakalov and Alexander Kirillov Jr. (2001). *Lectures on tensor categories and modular functors*. Vol. 21. University Lecture Series. American Mathematical Society, Providence, RI, pp. x+221. MR: 1797619 (cit. on p. 1313).

Vladimir Baranovsky (2015). "Uhlenbeck compactification as a functor". *Int. Math. Res. Not. IMRN* 23, pp. 12678–12712. MR: 3431633 (cit. on p. 1308).

Alexander Beilinson and Vladimir Drinfeld (2000). *Quantization of Hitchin's integrable system and Hecke eigensheaves* (cit. on pp. 1301, 1302).

Tom Braden (2003). "Hyperbolic localization of intersection cohomology". *Transform. Groups* 8.3, pp. 209–216. MR: 1996415 (cit. on p. 1303).

Alexander Braverman, Pavel Etingof, and Michael Finkelberg (Nov. 2016). "Cyclotomic double affine Hecke algebras (with an appendix by Hiraku Nakajima and Daisuke Yamakawa)". arXiv: 1611.10216 (cit. on p. 1317).

Alexander Braverman and Michael Finkelberg (2010). "Pursuing the double affine Grassmannian. I. Transversal slices via instantons on A_k-singularities". *Duke Math. J.* 152.2, pp. 175–206. MR: 2656088 (cit. on pp. 1306, 1307).

– (2012). "Pursuing the double affine Grassmannian II: Convolution". *Adv. Math.* 230.1, pp. 414–432. MR: 2900549 (cit. on pp. 1306–1308).

– (2013). "Pursuing the double affine Grassmannian III: Convolution with affine zastava". *Mosc. Math. J.* 13.2, pp. 233–265, 363. MR: 3134906 (cit. on pp. 1306–1308).

– (2014). "Semi-infinite Schubert varieties and quantum K-theory of flag manifolds". *J. Amer. Math. Soc.* 27.4, pp. 1147–1168. MR: 3230820 (cit. on pp. 1315, 1316).

Alexander Braverman, Michael Finkelberg, and Dennis Gaitsgory (2006). "Uhlenbeck spaces via affine Lie algebras". In: *The unity of mathematics*. Vol. 244. Progr. Math. see http://arxiv.org/abs/math/0301176 for erratum. Birkhäuser Boston, Boston, MA, pp. 17–135. MR: 2181803 (cit. on pp. 1307, 1308, 1315, 1316).

Alexander Braverman, Michael Finkelberg, Dennis Gaitsgory, and Ivan Mirković (2002). "Intersection cohomology of Drinfeld's compactifications". *Selecta Math. (N.S.)* 8.3. (See erratum in Braverman, Finkelberg, Gaitsgory and Mirković [2004]), pp. 381–418. MR: 1931170 (cit. on p. 1314).

– (2004). "Erratum to: Intersection cohomology of Drinfeld's compactifications". *Selecta Math. (N.S.)* 10. (Erratum to Braverman, Finkelberg, Gaitsgory and Mirković [2002]), pp. 429–430. arXiv: 0012129. MR: 2099075.

Alexander Braverman, Michael Finkelberg, and Hiraku Nakajima (Apr. 2016a). "Coulomb branches of $3d\,\mathcal{N} = 4$ quiver gauge theories and slices in the affine Grassmannian (with appendices by Alexander Braverman, Michael Finkelberg, Joel Kamnitzer, Ryosuke Kodera, Hiraku Nakajima, Ben Webster, and Alex Weekes)". arXiv: 1604.03625 (cit. on pp. 1304, 1308, 1311, 1312, 1314–1316).

– (2016b). "Instanton moduli spaces and $\mathcal{W}$-algebras". *Astérisque* 385, pp. vii+128. MR: 3592485 (cit. on p. 1308).

– (Jan. 2016c). "Towards a mathematical definition of Coulomb branches of 3-dimensional $\mathcal{N} = 4$ gauge theories, II". arXiv: 1601.03586 (cit. on p. 1308).

– (June 2017). "Ring objects in the equivariant derived Satake category arising from Coulomb branches (with an appendix by Gus Lonergan)". arXiv: 1706.02112 (cit. on p. 1308).

Alexander Braverman and Dennis Gaitsgory (2001). "Crystals via the affine Grassmannian". *Duke Math. J.* 107.3, pp. 561–575. MR: 1828302 (cit. on p. 1304).

Alexander Braverman and David Kazhdan (2013). "Representations of affine Kac-Moody groups over local and global fields: a survey of some recent results". In: *European Congress of Mathematics*. Eur. Math. Soc., Zürich, pp. 91–117. MR: 3469117 (cit. on p. 1306).

Alexander Braverman, David Kazhdan, and Manish M. Patnaik (2016). "Iwahori-Hecke algebras for p-adic loop groups". *Invent. Math.* 204.2, pp. 347–442. MR: 3489701 (cit. on p. 1306).

Stefano Cremonesi, Amihay Hanany, and Alberto Zaffaroni (2014). "Monopole operators and Hilbert series of Coulomb branches of $3d\,\mathcal{N} = 4$ gauge theories". *Journal of High Energy Physics* 1401, p. 005 (cit. on p. 1313).

Stefano Cremonesi, Noppadol Mekareeya, Amihay Hanany, and Giulia Ferlito (2014). "Coulomb branch and the moduli space of instantons". *Journal of High Energy Physics* 1412.arXiv: 1408.6835, p. 103 (cit. on p. 1314).

Vladimir Drinfeld and Dennis Gaitsgory (2014). "On a theorem of Braden". *Transform. Groups* 19.2, pp. 313–358. MR: 3200429 (cit. on p. 1303).

Boris Feigin, Evgeny Feigin, Michio Jimbo, Tetsuji Miwa, and Evgeny Mukhin (2009). "Fermionic formulas for eigenfunctions of the difference Toda Hamiltonian". *Lett. Math. Phys.* 88.1-3, pp. 39–77. MR: 2512140 (cit. on p. 1315).

Evgeny Feigin and Ievgen Makedonskyi (Sept. 2017). "Semi-infinite Plücker relations and Weyl modules". arXiv: 1709.05674 (cit. on p. 1315).

Michael Finkelberg, Victor Ginzburg, Andrei Ionov, and Alexander Kuznetsov (2016). "Intersection cohomology of the Uhlenbeck compactification of the Calogero-Moser space". *Selecta Math. (N.S.)* 22.4, pp. 2491–2534. MR: 3573963 (cit. on p. 1316).

Michael Finkelberg, Joel Kamnitzer, Khoa Pham, Leonid Rybnikov, and Alex Weekes (2018). "Comultiplication for shifted Yangians and quantum open Toda lattice". *Adv. Math.* 327, pp. 349–389. arXiv: 1608.03331. MR: 3761996 (cit. on p. 1316).

Michael Finkelberg and Dmitry V. Kubrak (2015). "Vanishing cycles on Poisson varieties". *Funct. Anal. Appl.* 49.2. Translation of Funktsional. Anal. i Prilozhen. **4**9 (2015), no. 2, 70–78, pp. 135–141. MR: 3374904 (cit. on p. 1308).

Michael Finkelberg and Ivan Mirković (1999). "Semi-infinite flags. I. Case of global curve $\mathbf{P}^1$". In: *Differential topology, infinite-dimensional Lie algebras, and applications*. Vol. 194. Amer. Math. Soc. Transl. Ser. 2. Amer. Math. Soc., Providence, RI, pp. 81–112. MR: 1729360 (cit. on p. 1304).

Michael Finkelberg and Leonid Rybnikov (2014). "Quantization of Drinfeld zastava in type C". *Algebr. Geom.* 1.2, pp. 166–180. MR: 3238110 (cit. on p. 1316).

Igor B. Frenkel (1982). "Representations of affine Lie algebras, Hecke modular forms and Korteweg-de Vries type equations". In: *Lie algebras and related topics (New Brunswick, N.J., 1981)*. Vol. 933. Lecture Notes in Math. Springer, Berlin-New York, pp. 71–110. MR: 675108 (cit. on p. 1307).

Stéphane Gaussent and Guy Rousseau (2014). "Spherical Hecke algebras for Kac-Moody groups over local fields". *Ann. of Math. (2)* 180.3, pp. 1051–1087. MR: 3245012 (cit. on p. 1306).

Victor Ginzburg (Nov. 1995). "Perverse sheaves on a Loop group and Langlands' duality". arXiv: alg-geom/9511007 (cit. on pp. 1301, 1302).

Victor Ginzburg and Simon Riche (2015). "Differential operators on G/U and the affine Grassmannian". *J. Inst. Math. Jussieu* 14.3, pp. 493–575. MR: 3352528 (cit. on p. 1303).

Alexander Grothendieck (1957). "Sur la classification des fibrés holomorphes sur la sphère de Riemann". *Amer. J. Math.* 79, pp. 121–138. MR: 0087176 (cit. on p. 1303).

Nicolas Guay, Hiraku Nakajima, and Curtis Wendlandt (Jan. 2017). "Coproduct for the Yangian of an affine Kac-Moody algebra". arXiv: 1701.05288 (cit. on p. 1317).

Tatsuyuki Hikita (2017). "An algebro-geometric realization of the cohomology ring of Hilbert scheme of points in the affine plane". *Int. Math. Res. Not. IMRN* 8, pp. 2538–2561. MR: 3658207 (cit. on p. 1313).

Joel Kamnitzer, Peter Tingley, Ben Webster, Alex Weekes, and Oded Yacobi (Nov. 2015). "Highest weights for truncated shifted Yangians and product monomial crystals". arXiv: 1511.09131 (cit. on p. 1313).

Joel Kamnitzer, Ben Webster, Alex Weekes, and Oded Yacobi (2014). "Yangians and quantizations of slices in the affine Grassmannian". *Algebra Number Theory* 8.4, pp. 857–893. MR: 3248988 (cit. on p. 1303).

Masaki Kashiwara (1989). "The flag manifold of Kac-Moody Lie algebra". In: *Algebraic analysis, geometry, and number theory (Baltimore, MD, 1988)*. Johns Hopkins Univ. Press, Baltimore, MD, pp. 161–190. MR: 1463702 (cit. on p. 1303).

– (1995). "On crystal bases". In: *Representations of groups (Banff, AB, 1994)*. Vol. 16. CMS Conf. Proc. Amer. Math. Soc., Providence, RI, pp. 155–197. MR: 1357199 (cit. on p. 1304).

Masaki Kashiwara and Toshiyuki Tanisaki (1995). "Kazhdan-Lusztig conjecture for affine Lie algebras with negative level". *Duke Math. J.* 77.1, pp. 21–62. MR: 1317626 (cit. on p. 1303).

Vasily Krylov (Sept. 2017). "Integrable crystals and restriction to Levi via generalized slices in the affine Grassmannian". arXiv: 1709.00391 (cit. on pp. 1304, 1312).

George Lusztig (1983). "Singularities, character formulas, and a q-analog of weight multiplicities". In: *Analysis and topology on singular spaces, II, III (Luminy, 1981)*. Vol. 101. Astérisque. Soc. Math. France, Paris, pp. 208–229. MR: 737932 (cit. on pp. 1301, 1302).

Ivan Mirković and Kari Vilonen (2007). "Geometric Langlands duality and representations of algebraic groups over commutative rings". *Ann. of Math. (2)* 166.1, pp. 95–143. MR: 2342692 (cit. on pp. 1301, 1302).

Gregory Moore and Nathan Seiberg (1989). "Classical and quantum conformal field theory". *Comm. Math. Phys.* 123.2, pp. 177–254. MR: 1002038 (cit. on p. 1313).

Hiraku Nakajima (2002). "Geometric construction of representations of affine algebras". In: *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*. Higher Ed. Press, Beijing, pp. 423–438. MR: 1989196 (cit. on p. 1307).

– (2009). "Quiver varieties and branching". *SIGMA Symmetry Integrability Geom. Methods Appl.* 5, Paper 003, 37. MR: 2470410 (cit. on pp. 1307, 1312).

– (Oct. 2015). "Questions on provisional Coulomb branches of 3-dimensional $\mathcal{N} = 4$ gauge theories". arXiv: 1510.03908 (cit. on p. 1308).

– (2016a). "Towards a mathematical definition of Coulomb branches of 3-dimensional $\mathcal{N} = 4$ gauge theories, I". *Adv. Theor. Math. Phys.* 20.3, pp. 595–669. MR: 3565863 (cit. on p. 1308).

– (2016b). "Towards a mathematical definition of Coulomb branches of 3-dimensional $\mathcal{N} = 4$ gauge theories, I". *Adv. Theor. Math. Phys.* 20.3, pp. 595–669. arXiv: 1706.05154. MR: 3565863 (cit. on p. 1308).

Hiraku Nakajima and Yuuya Takayama (2017). "Cherkis bow varieties and Coulomb branches of quiver gauge theories of affine type A". *Selecta Math. (N.S.)* 23.4, pp. 2553–2633. MR: 3703461 (cit. on pp. 1311, 1314).

Raphaël Rouquier (2012). "Quiver Hecke algebras and 2-Lie algebras". *Algebra Colloq.* 19.2, pp. 359–410. MR: 2908731 (cit. on p. 1317).

Nathan Seiberg and Edward Witten (1997). "Gauge dynamics and compactification to three dimensions". In: *The mathematical beauty of physics (Saclay, 1996)*. Vol. 24. Adv. Ser. Math. Phys. World Sci. Publ., River Edge, NJ, pp. 333–366. MR: 1490862 (cit. on p. 1309).

William Slofstra (2012). "A Brylinski filtration for affine Kac-Moody algebras". *Adv. Math.* 229.2, pp. 968–983. MR: 2855083 (cit. on p. 1308).

Ben Webster (Sept. 2016). "Representation theory of the cyclotomic Cherednik algebra via the Dunkl-Opdam subalgebra". arXiv: 1609.05494 (cit. on p. 1317).

Edward Witten (2010). "Geometric Langlands from six dimensions". In: *A celebration of the mathematical legacy of Raoul Bott*. Vol. 50. CRM Proc. Lecture Notes. Amer. Math. Soc., Providence, RI, pp. 281–310. MR: 2648898 (cit. on p. 1308).

Received 2017-12-17.

MICHAEL FINKELBERG
NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS
DEPARTMENT OF MATHEMATICS
6 USACHEVA ST.
MOSCOW 119048

and

SKOLKOVO INSTITUTE OF SCIENCE AND TECHNOLOGY
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS
fnklberg@gmail.com

PROC. INT. CONG. OF MATH. – 2018
Rio de Janeiro, Vol. 2 (1321–1338)

REPRESENTATIONS OF GALOIS ALGEBRAS

VYACHESLAV FUTORNY

Abstract

Galois algebras allow an effective study of their representation theory based on the invariant skew group structure. In particular, this leads to many remarkable results on Gelfand-Tsetlin representations of the general linear Lie algebra $\mathfrak{gl}_n$, quantum $\mathfrak{gl}_n$, Yangians of type A and finite W-algebras of type A.

1 Introduction

A classical problem of the representation theory of simple complex finite dimensional Lie algebras is the classification of simple modules. Today such a classification is known only for the Lie algebra $\mathfrak{sl}_2$ Block [1981]. Special attention is addressed to the study of so-called *weight* modules, i.e. those on which a certain Cartan subalgebra is diagonalizable. By the results of Fernando [1990] and Mathieu [2000], the classification of simple weight modules with finite dimensional weight subspaces is well known for any simple finite dimensional Lie algebra. On the other hand, a classification of simple modules remains open even in the category of weight modules with infinite dimensional weight subspaces. The largest subcategory of the category of weight module with some understanding of simple objects is the category of Gelfand-Tsetlin modules. The Gelfand-Tsetlin theory has attracted considerable interest in the last 40 years after the pioneering work of Gelfand and Cetlin [1950] and was developed in Drozd, Ovsienko, and Futorny [1991], Graev [2004], Graev [2007], Drozd, Ovsienko, and Futorny [1989], Mazorchuk [1998], Mazorchuk [2001], Molev [2006], Želobenko [1973], among the others. Gelfand-Tsetlin integrable systems were studied by Guillemin and Sternberg [1983], Kostant and Wallach [2006a], Kostant and Wallach [2006b], Colarusso and Evens [2010], Colarusso and Evens [2014].

Gelfand-Tsetlin theory can be viewed in a more general context of Harish-Chandra categories Drozd, Futorny, and Ovsienko [1994], Futorny and Ovsienko [2007] which play very important role in the representation theory. These are the categories of modules over

MSC2010: primary 17B10; secondary 17B05, 16G30.

a given algebra defined by the restriction onto a fixed subalgebra. General setting for the study of Harish-Chandra categories was established in Drozd, Futorny, and Ovsienko [1994]. Examples of Harish-Chandra categories include classical Harish-Chandra modules over a finite dimensional Lie algebra defined with respect to a reductive subalgebra (Dixmier [1974]), weight modules over semisimple finite dimensional Lie algebras with respect to a Cartan subalgebra, Gelfand-Tsetlin modules over $\mathfrak{gl}_n$ (Drozd, Ovsienko, and Futorny [1991]), certain representations of Yangians (Futorny, Molev, and Ovsienko [2005]) etc. In the case of generalized Weyl algebras of rank 1 this approach led to a complete classification of simple modules (Bavula [1992], Bavula and van Oystaeyen [2004]).

Developed techniques turned out to be very useful in the study of Gelfand-Tsetlin modules for the Lie algebra $\mathfrak{gl}_n$ (Drozd, Ovsienko, and Futorny [1991], Ovsienko [2002]). Gelfand-Tsetlin modules form the full subcategory of weight $\mathfrak{gl}_n$-modules which are sums of finite dimensional modules over the *Gelfand-Tsetlin subalgebra* Γ (certain maximal commutative subalgebra of the universal enveloping algebra of $\mathfrak{gl}_n$) Drozd, Ovsienko, and Futorny [1991], Futorny and Ovsienko [2010]. These modules are weight modules with respect to some Cartan subalgebra of $\mathfrak{gl}_n$ but they allow to have infinite dimensional weight spaces.

Gelfand-Tsetlin theory had a successful development in Ovsienko [2002], where it was shows that simple Gelfand-Tsetlin modules over $\mathfrak{gl}_n$ are parametrized up to some finiteness by the maximal ideals of Γ. Different explicit constructions of Gelfand-Tsetlin modules for $\mathfrak{gl}_n$ were recently obtained in Futorny, Grantcharov, and Ramirez [2014], Futorny, Grantcharov, and Ramirez [2015], Futorny, Grantcharov, and Ramirez [2016b], Futorny, Grantcharov, and Ramirez [2016a], Futorny, Ramirez, and Zhang [2016], Zadunaisky [2017], Vishnyakova [2018], Vishnyakova [2017], Ramírez and Zadunaisky [2017]. Nevertheless, the problem remains open.

As an attempt to unify the representation theories of the universal enveloping algebra of $\mathfrak{gl}_n$ and of the generalized Weyl algebras a new concept of Galois orders was introduced in Futorny and Ovsienko [2010]. These algebras have a hidden skew (semi)group structure. In particular, the universal enveloping algebra of $\mathfrak{gl}_n$ is an example of such algebra where invariant skew group structure comes from the Gelfand-Tsetlin formulas. Representation theory of Galois algebras was developed in Futorny and Ovsienko [2014]. It provides a new framework for the study of representation of various classes of algebras. Recent paper of Hartwig (Hartwig [2017a]) discovers new examples of Galois algebras for which the theory can be effectively applied.

2 Harish-Chandra modules

We recall basic facts about Harish-Chandra module categories following Drozd, Futorny, and Ovsienko [1994]. Let U be an associative algebra over $\mathbb{k}$, $\Gamma \subset U$ a subalgebra. The set of maximal ideals $\mathbf{m}$ of Γ such that $\dim \Gamma/\mathbf{m} < \infty$ will be called the *cofinite spectrum* $\operatorname{cfs}\Gamma$ of Γ. Then every $\mathbf{m} \in \operatorname{cfs}\Gamma$ defines a unique simple Γ-module of dimension $l(\mathbf{m})$ where $\Gamma/\mathbf{m} \simeq M_{l(\mathbf{m})}(\mathbb{k})$.

We say that M is a *Harish-Chandra module* (with respect to Γ) if M a finitely generated U-module such that

$$M = \bigoplus_{\mathbf{m}\in \operatorname{cfs}\Gamma} M(\mathbf{m}),$$

where

$$M(\mathbf{m}) = \{x \in M \mid \exists k,\ \mathbf{m}^k x = 0\}.$$

The support of a Harish-Chandra module M is a subset of $\operatorname{cfs}\Gamma$ consisting of those $\mathbf{m}$ for which $M(\mathbf{m}) \neq 0$.

We denote by $\mathbb{H}(U,\Gamma)$ the full subcategory consisting of all Harish-Chandra modules in $U -$ mod. It is closed under the operations of taking submodules, quotients and direct sums.

In Drozd, Futorny, and Ovsienko [ibid.] a concept of a Harish-Chandra subalgebra was introduced. For any $\mathbf{m} \in \operatorname{cfs}\Gamma$ denote by $L_{\mathbf{m}}$ the unique simple $\Gamma/\mathbf{m}$-module. We say that Γ is quasi-commutative if $\operatorname{Ext}^1(L_{\mathbf{m}}, L_{\mathbf{n}}) = 0$ for all $\mathbf{m} \neq \mathbf{n}$. We also say that Γ is quasi-central if for every $u \in U$, the Γ-bimodule $\Gamma u \Gamma$ is finitely generated as a left and as a right Γ-module. Clearly, for a noetherian Γ it is sufficiently to check this condition only for the generators of Γ (cf. Drozd, Futorny, and Ovsienko [ibid.], Proposition 8). A subalgebra Γ is called *Harish-Chandra* if it is quasi-central and quasi-commutative.

Example 2.1. *Let $\mathfrak{g}$ be a finite dimensional Lie algebra and $\mathfrak{F}$ its reductive Lie subalgebra. Then $\Gamma = U(\mathfrak{F})$ is a Harish-Chandra subalgebra of $U = U(\mathfrak{g})$. Indeed, Γ is quasi-commutative since any $\mathbf{m} \in \operatorname{cfs}\Gamma$ is cofinite. Also Γ is quasi-central since $\mathfrak{g}$ is finite dimensional.*

The concept of a Harish-Chandra subalgebra is essential for understanding the categories of Harish-Chandra modules. We address to Drozd, Futorny, and Ovsienko [ibid.] for further properties of Harish-Chandra subalgebras in the general setting. The most studied is the case of commutative Γ which we consider next.

2.1 Commutative Γ. Let U be an associative algebra and $\Gamma \subset U$ a noetherian commutative subalgebra. A natural idea is to try to parametrize simple modules in the Harish-Chandra category $\mathbb{H}(U,\Gamma)$ by simple Γ-modules. Since any simple Γ-module is 1-dimensional it defines a homomorphism from Γ to $\Bbbk$ which we call a *character* of Γ. The kernel of any such character is a maximal ideal of Γ, thus there is a one-to-one correspondence between the characters of Γ and elements of $\operatorname{Specm}\Gamma$. The following problem of extension of characters to simple modules in $\mathbb{H}(U,\Gamma)$ is of prime importance for description of possible supports of simple Harish-Chandra modules.

Problem 1. Given $\mathbf{m} \in \operatorname{Specm}\Gamma$ is there a (simple) module $M \in \mathbb{H}(U,\Gamma)$ such that $M(\mathbf{m}) \neq 0$.

Recall that in the classical setting when both U and Γ are commutative and extension $\Gamma \subset U$ is integral then we have a map between the sets of prime (maximal) ideals $\varphi : \operatorname{Spec} U \to \operatorname{Spec}\Gamma$ and the fiber $\varphi^{-1}(\mathfrak{p})$ is non-empty for every $\mathfrak{p} \in \operatorname{Spec} A$. In particular, every character of Γ can be extended to a character of U. Moreover, if U is finitely generated module over Γ then all fibers $\varphi^{-1}(\mathfrak{p})$ are finite and the number of different extensions of each character of Γ is finite. By the Hilbert-Noether theorem this is the case when $U = S(V)$ is the symmetric algebra of a finite dimensional vector space V and Γ is the subalgebra of G-invariants of U for some finite subgroup G of $GL(V)$.

If U is noncommutative then the restriction functor from the category $\mathbb{H}(U,\Gamma)$ to the category of torsion Γ-modules induces a map Φ from $\operatorname{Specm}\Gamma$ to the set of isomorphism classes $Irr(U)$ of simple U-modules in $\mathbb{H}(U,\Gamma)$. Given a maximal ideal $\mathbf{m} \in \operatorname{Specm}\Gamma$, $\Phi(\mathbf{m})$ consist of those simple $V \in \mathbb{H}(U,\Gamma)$ such that $V(\mathbf{m}) \neq 0$ (or left maximal ideals of U which contain $\mathbf{m}$).

Example 2.2. *(i) Let $\mathfrak{g}$ be a reductive Lie algebra with a Cartan subalgebra $\mathfrak{H}$, $U = U(\mathfrak{g})$ and $\Gamma = U(\mathfrak{H})$. Then for any weight $\lambda \in \mathfrak{H}^*$ the fiber $\Phi(\lambda)$ is infinite (even in the category $\mathcal{O}$ which is a subcategory of $\mathbb{H}(U,\Gamma)$).*

(ii) Let $U = U_n = \mathbb{C}[S_n]$, $U_1 \subset \ldots \subset U_n$, Z_k the center of $\mathbb{C}[U_k]$. Then $\Gamma =< Z_1,\ldots,Z_n >$ is maximal commutative. It is generated by the Jucys-Murphy elements $X_i = (1i)+\ldots+(i-1i)$, $i = 1,\ldots,n$. The elements of $\operatorname{Specm}\Gamma$ *parametrize the irreducible representations of the group S_n Okounkov and Vershik [1996].*

The freeness of U over Γ as a right module guarantees the lifting of characters of Γ (all examples above are of this kind). Finding sufficient conditions for the fiber $\Phi(\mathbf{m})$ to be non-empty for any point $\mathbf{m} \in \operatorname{Specm}\Gamma$ is a difficult problem in general. In particular, if U is a *special filtered* such conditions were obtained in Futorny and Ovsienko [2005] generalizing the Kostant's theorem (see further examples in Futorny and Ovsienko [ibid.] and Futorny, Molev, and Ovsienko [2005]).

3 Galois algebras

A class of Galois rings (orders) was introduced in Futorny and Ovsienko [2010] to deal with the problem of extension of characters of commutative subalgebras.

Let R be a ring, $\mathcal{M}$ a monoid acting on R by ring automorphisms. We will denote the action of $m \in \mathcal{M}$ on $r \in R$ by r^m. Consider the skew monoid ring $R * \mathcal{M}$. Any element of $R * \mathcal{M}$ can be written in the form $x = \sum_{m \in \mathcal{M}} x_m m$. Define $supp\, x$ as the set of those $m \in \mathcal{M}$ for which x_m is not zero.

Let G be a finite group acting on $\mathcal{M}$ by conjugation. Then we have an action of G on $R * \mathcal{M}$ by ring automorphisms: $g(rm) = g(r)g(m)$, $g \in G, r \in R, m \in \mathcal{M}$.

Assume now that Γ is an integral domain, K the field of fractions of Γ and L a finite Galois extension of K with the Galois group $G = \mathrm{Gal}(L/K)$. Consider the action of G by conjugation on $\mathrm{Aut}(L)$. Let $\mathcal{M}$ be any G-invariant submonoid of $\mathrm{Aut}(L)$. For our purposes we will always require $\mathcal{M}$ to be K-*separating*, that is $m_1|_K = m_2|_K \Rightarrow m_1 = m_2$ for $m_1, m_2 \in \mathcal{M}$. The action of G on L and on $\mathcal{M}$ (by conjugations) extends to the action of G on the skew monoid ring $L * \mathcal{M}$. Denote by $\mathcal{K} = (L * \mathcal{M})^G$ the subring of invariants.

Definition 3.1. *A finitely generated* Γ*-subring* U *of* $\mathcal{K}$ *is called a* Galois ring over Γ *if* $UK = KU = \mathcal{K}$.

We have the following characterization of Galois rings:

We will always assume that all Galois rings are $\Bbbk$-algebras. In this case we say that a Galois ring is a Galois algebra over Γ.

Example 3.1. *Let* $U = \Gamma(\sigma, a)$ *be a* generalized Weyl algebra *of rank* 1 *(Bavula [1992]), where* Γ *is a unital integral domain,* $a \in \Gamma$*,* σ *an automorphism of* Γ *of infinite order. It is generated over* Γ *by* X *and* Y *such that* $X\gamma = \sigma(\gamma)X$*,* $Y\gamma = \sigma^{-1}(\gamma)Y$*,* $YX = a$*,* $XY = \sigma(a)$*. Let* K *be the field of fractions of* Γ *and* $\mathcal{M} \simeq \mathbb{Z}$ *is a subgroup of* $\mathrm{Aut}\,\Gamma$ *generated by* σ*. Then* U *can be embedded into the skew group algebra* $K * \mathbb{Z}$ *when* $X \mapsto \sigma$ *and* $Y \mapsto a\sigma^{-1}$*. Clearly,* U *is a Galois algebra over* Γ*. Note that* $U \simeq \Gamma * \mathbb{Z}$ *if* a *is invertible in* Γ*.*

3.1 Galois orders.

Now we discuss a special class of Galois rings which are called *Galois orders*. Galois orders were introduced in Futorny and Ovsienko [2010] as a natural noncommutative generalization of a classical notion of order in skew group rings (cf. McConnell and Robson [1987]).

A Galois ring U over Γ. is *right (respectively left) Galois order*, if for any finite dimensional right (respectively left) K-subspace $W \subset U[S^{-1}]$ (respectively $W \subset [S^{-1}]U$), $W \cap U$ is a finitely generated right (respectively left) Γ-module. A Galois ring is *Galois order* if it is both right and left Galois order.

For a right Γ-submodule $M \subset U$ denote

$$\mathbb{D}_r(M) = \{u \in U \mid \exists \gamma \in \Gamma, \gamma \neq 0 \text{ such that } u \cdot \gamma \in M\}.$$

It follows immediately that $\mathbb{D}_r(M)$ is a Γ-module.

We have the following characterization of a Galois order.

Proposition 3.1 (Futorny and Ovsienko [2010], Corollary 5.1, 5.2)**.** *(i) A Galois ring U over a noetherian Γ is* right (left) Galois order *if and only if for every finitely generated right (left) Γ-module $M \subset U$, the right (left) Γ-module $\mathbb{D}_r(M)$ is finitely generated.*

(ii) If a Galois ring U over a noetherian domain Γ is projective as a right (left) Γ-module then U is a right (left) Galois order.

In the commutative case if K is the field of fractions of Γ, $U \subset K$ is finitely generated over Γ and the extension $\Gamma \subset U$ is integral then U is Galois order over Γ. Further examples of Galois orders include: generalized Weyl algebras over integral domains with infinite order automorphisms (e.g. the n-th Weyl algebra A_n, the quantum plane, the q-deformed Heisenberg algebra, quantized Weyl algebras, the Witten-Woronowicz algebra Bavula [1992]; the universal enveloping algebra of $\mathfrak{gl}_n$ over the Gelfand-Tsetlin subalgebra Drozd, Ovsienko, and Futorny [1991], Drozd, Futorny, and Ovsienko [1994], finite W-algebras Futorny, Molev, and Ovsienko [2005]).

There is a strong connection between Galois orders and maximality of Harish-Chandra subalgebras. Namely, we have

Theorem 3.1. *(i) Let Γ be a finitely generated domain over $\mathbb{k}$ and U a Galois order over Γ. Then Γ is a Harish-Chandra subalgebra in U.*

(ii) Let U be a Galois ring over finitely generated $\mathbb{k}$-algebra Γ and $\mathcal{M}$ be a group. If Γ is a Harish-Chandra subalgebra in U then U is a Galois order if and only if U_e is an integral extension of Γ.

(iii) Let U be a Galois ring over a normal noetherian Harish-Chandra subalgebra Γ and $\mathcal{M}$ be a group. Then U is a Galois order over Γ if and only if Γ is maximal commutative in U.

Proof. First item follows from [Futorny and Ovsienko [2010], Corollary 5.4]. Second item follows from [Futorny and Ovsienko [ibid.], Theorem 5.2, (2)]. Let U be a Galois ring over a normal noetherian Harish-Chandra subalgebra Γ. If Γ is maximal then applying [Futorny and Ovsienko [ibid.], Corollary 5.6, (2)] and the fact that U has no torsion as a Γ-module we conclude that U is a Galois order over Γ. To prove the converse, it is sufficient to show that $U \cap K = \Gamma$ by [Futorny and Ovsienko [ibid.], Theorem 4.1]. Since $U \cap K$ is an integral extension of Γ by the second item, the statement follows from the normality of Γ.

□

The problem of lifting of characters for Galois orders was studied in Futorny and Ovsienko [2014]. In particular, sufficient conditions for the fiber $\Phi(\mathbf{m})$ to be nontrivial and finite were established. Let $U \subset (L * \mathcal{M})^G$ be a Galois ring over Γ. Consider the integral closure $\bar{\Gamma}$ of Γ in L. It is a standard fact that if Γ is finitely generated as a $\Bbbk$-algebra then any character of Γ has finitely many extensions to characters of $\bar{\Gamma}$.

Let $\bar{\mathbf{m}}$ be any lifting of $\mathbf{m}$ to the integral closure of Γ in L, and $\mathcal{M}_{\mathbf{m}}$ the stabilizer of $\bar{\mathbf{m}}$ in $\mathcal{M}$. Note that the group $\mathcal{M}_{\mathbf{m}}$ is defined uniquely up to G-conjugation. Thus the cardinality of $\mathcal{M}_{\mathbf{m}}$ does not depend on the choice of the lifting. We denote it by $|\mathbf{m}|$.

For $\mathbf{m}, \mathbf{n} \in \operatorname{Specm} \Gamma$ set

$$S(\mathbf{m}, \mathbf{n}) = \{m \in \mathcal{M} \mid \bar{\mathbf{n}} \in GmG \cdot \bar{\mathbf{m}}\},$$

which is a G-invariant subset in $\mathcal{M}$. If $\mathcal{M}$ is a group then we have

$$|S(\mathbf{m}, \mathbf{n})/G| \leq |\{x \in \mathcal{M} \mid x\bar{\mathbf{m}} = \bar{\mathbf{n}}\}|.$$

Denote by $r(\mathbf{m}, \mathbf{n})$ the minimal number of generators of $U(S(\mathbf{m}, \mathbf{n}))$ as a right Γ-module.

Theorem 3.2. *[Futorny and Ovsienko [ibid.], Theorem A, , Theorem 8] Let Γ be a commutative domain which is finitely generated as a $\Bbbk$-algebra, $U \subset (L * \mathcal{M})^G$ a right Galois order over Γ, $\mathbf{m} \in \operatorname{Specm} \Gamma$. Suppose $|\mathbf{m}|$ is finite.*

(i) The fiber $\Phi(\mathbf{m})$ is non-empty.

(ii) If U is a Galois order over Γ, then the fiber $\Phi(\mathbf{m})$ is finite.

(iii) Let U be a Galois order over normal noetherian Γ, $M \in \mathbb{H}(U, \Gamma)$ a simple U-module and $\mathbf{m} \in \operatorname{Specm} \Gamma$. If U is free as a right Γ-module then for any $\mathbf{n}$

$$\dim_{\Bbbk} M(\mathbf{n}) \leq |S(\mathbf{m}, \mathbf{n})/G|.$$

3.2 Principal Galois orders. Further examples of Galois orders were recently constructed by Hartwig [2017a]. As before denote $\mathcal{K} = (L * \mathcal{M})^G$. Clearly, $\mathcal{K}$ is its own Galois subring over Γ. But, $\mathcal{K}$ is a Galois order if and only if $\Lambda = L$, where Λ is the integral closure of Γ in L [Hartwig [ibid.], Corollary 2.15].

Let $x = \sum_{\phi \in \mathcal{M}} x_\phi \phi \in L * \mathcal{M}$ and $a \in L$. Define the evaluation $x(a) := \sum_{\phi \in \mathcal{M}} x_\phi \phi(a)$ [Hartwig [ibid.], Definition 2.18]. Then we have

Theorem 3.3 (Hartwig [2017a], Theorem 2.21)**.**

$$\mathcal{K}_\Gamma = \{x \in \mathcal{K} \mid x(\gamma) \in \Gamma \textit{ for all } \gamma \in \Gamma\}$$

is a Galois order over Γ *in* $\mathcal{K}$.

One immediately sees that any Galois subring of $\mathcal{K}_\Gamma$ over Γ is a Galois order. Such orders are called *principal* Hartwig [ibid.].

A new class of principal Galois orders, *rational Galois orders*, was introduced in Hartwig [ibid.]. These structures are attached to an arbitrary finite reflection group and a set of difference operators with rational function coefficients. In particular, the parabolic subalgebras of finite W-algebras of type A are rational Galois orders [Hartwig [ibid.], Theorem 1.2]. This extends the result of Futorny, Molev, and Ovsienko [2005] for W-algebras of type A. Other examples of principal Galois orders include *orthogonal Gelfand-Tsetlin algebras* (Hartwig [2017a], Theorem 4.6) introduced in Mazorchuk [1999] and *quantum orthgonal Gelfand-Tsetlin algebras* (Hartwig [2017a], Theorem 5.6) introduced in Hartwig [2017b]. The family of quantum orthogonal Gelfand-Tsetlin algebras includes in particular quantized universal enveloping algebra $U_q(\mathfrak{gl}_n)$ and, as a consequence, implies the maximality of the Gelfand-Tsetlin subalgebra of $U_q(\mathfrak{gl}_n)$ when q is not a root of unity (this was conjectured by Mazorchuk and Turowska [2000]).

4 Gelfand-Tsetlin modules

Now we address the Lie algebra $\mathfrak{gl}_n$ consisting of all $n \times n$ complex matrices with the standard basis of elementary matrices $\{e_{i,j} \mid 1 \leq i, j \leq n\}$. For each $k \leqslant n$ denote by $\mathfrak{gl}_k$ the Lie subalgebra of $\mathfrak{gl}_n$ spanned by $\{e_{ij} \mid i, j = 1, \dots, k\}$. We have the following embeddings of Lie subalgebras

$$\mathfrak{gl}_1 \subset \mathfrak{gl}_2 \subset \dots \subset \mathfrak{gl}_n.$$

We have corresponding embeddings $U_1 \subset U_2 \subset \dots \subset U_n$ of the universal enveloping algebras $U_k = U(\mathfrak{gl}_k)$, $1 \leq k \leq n$. Set $U = U_n$.

Let Z_k be the center of U_k. This is the polynomial algebra generated by the following elements

$$c_{ks} = \sum_{(i_1,\dots,i_s)\in\{1,\dots,k\}^s} e_{i_1 i_2} e_{i_2 i_3} \dots e_{i_s i_1}, \tag{1}$$

$s = 1, \dots, k$.

Let Γ be the subalgebra of $U(\mathfrak{gl}_n)$ generated by the centers Z_k, $k = 1, \dots, n$, the *Gelfand-Tsetlin subalgebra* Drozd, Ovsienko, and Futorny [1991]. The generators c_{ks}, $k = 1, \dots, n$, $s = 1, \dots, k$ are algebraically independent Želobenko [1973].

Let Λ be the polynomial algebra in the variables $\{\lambda_{ij} \mid 1 \leqslant j \leqslant i \leqslant n\}$. Consider the embedding $\pi : \Gamma \longrightarrow \Lambda$ such that

$$c_{ks} \mapsto \sum_{i=1}^{k} (\lambda_{ki} + k - 1)^s \prod_{j \neq i} \left(1 - \frac{1}{\lambda_{ki} - \lambda_{kj}}\right).$$

One can easily check that $\pi(c_{ks})$ is a symmetric polynomial in Λ of degree s in variables $\lambda_{k1}, \ldots, \lambda_{kk}$. Let $G = \prod_{i=1}^{n} S_i$ be the product of symmetric groups. Then G acts naturally on Λ where S_k permutes the variables $\lambda_{k1}, \ldots, \lambda_{kk}$, $k = 1, \ldots, n$. The image of Γ, $\pi(\Gamma)$, coincides with the subalgebra of G-invariant polynomials in Λ which we identify with Γ.

Consider the Harish-Chandra category $H(U, \Gamma)$. We will call the modules of $H(U, \Gamma)$ *Gelfand-Tsetlin modules*. If $M \in H(U, \Gamma)$ then

$$M = \bigoplus_{\mathsf{m} \in \operatorname{Specm} \Gamma} M(\mathbf{m}),$$

where $M(\mathsf{m}) = \{v \in M | \mathbf{m}^k v = 0 \text{ for some } k \geq 0\}$. Clearly, any simple Gelfand-Tsetlin module over $\mathfrak{gl}(n)$ is a weight module with respect to the Cartan subalgebra spanned by e_{ii}, $i = 1, \ldots, n$. Moreover, for $n = 2$ any simple weight module is a Gelfand-Tsetlin module. For $n > 2$ this is not true in general, but holds for modules with finite weight multiplicities. For a Gelfand-Tsetlin module $M(\mathbf{m}) \in H(U, \Gamma)$ and $\mathbf{m} \in \operatorname{Specm} \Gamma$ we call the dimension of $M(\mathbf{m})$ *the Gelfand-Tsetlin multiplicity* of $\mathbf{m}$.

4.1 Finite dimensional modules over $\mathfrak{gl}_n$. We recall a classical result of Gelfand and Cetlin [1950] which gives an explicit basis for all simple finite dimensional $\mathfrak{gl}_n$-modules.

For convenience we consider the elements of the space $\mathbb{C}^k$ as k-tuples whose entries are labeled as follows $(v_{k1,\ldots,v_{kk}})$. We also identify $\mathbb{C}^{\frac{n(n+1)}{2}}$ with $T_n(\mathbb{C}) = \mathbb{C}^n \times \mathbb{C}^{n-1} \times \ldots \times \mathbb{C}$. Then every vector v in $\mathbb{C}^{\frac{n(n+1)}{2}}$ can be written in the following form:

$$v = (v_{n1}, \ldots, v_{nn} | v_{n-1,1}, \ldots, v_{n-1,n-1} | \cdots | v_{21}, v_{22} | v_{11})$$

to which we associate the following array $T(v)$

$$\begin{array}{ccccccc}
v_{n1} & v_{n2} & & \cdots & & v_{n,n-1} & v_{nn} \\
& v_{n-1,1} & & \cdots & & v_{n-1,n-1} & \\
& & \cdots & \cdots & \cdots & & \\
& & v_{21} & & v_{22} & &
\end{array}$$

v_{11}

Such an array will be called a *Gelfand-Tsetlin tableau* of height n.

For a fixed element $v = (v_{ij})_{j\leq i=1}^{n} \in T_n(\mathbb{C})$ consider the set

$$v + T_{n-1}(\mathbb{Z}) = \{v + w \mid w = (w_{ij})_{j\leq i=1}^{n}, w_{ij} \in \mathbb{Z}, w_{nk} = 0\,, k = 1, \ldots, n\}.$$

Denote by $V(T(v))$ the complex vector space spanned by the set $v + T_{n-1}(\mathbb{Z})$ as a basis. Clearly, the spaces $V(T(v))$ and $T_n(\mathbb{C})$ are not isomorphic as $T(v+w) \neq T(v) + T(w)$ in $V(T(v))$.

A Gelfand-Tsetlin tableau $T(v)$ of height n is called *standard* if $v_{ki} - v_{k-1,i} \in \mathbb{Z}_{\geq 0}$ and $v_{k-1,i} - v_{k,i+1} \in \mathbb{Z}_{>0}$ for all $1 \leq i \leq k \leq n-1$.

Theorem 4.1 (Gelfand and Cetlin [1950]). *Let $L(\lambda)$ be the simple finite dimensional $\mathfrak{gl}_n$-module of highest weight $\lambda = (\lambda_1, \ldots, \lambda_n)$. Then the set of all standard tableaux $T(v)$ with fixed top row $v_{ni} = \lambda_i - i + 1$, $i = 1, \ldots, n$ forms a basis of $L(\lambda)$. Moreover, the action of the generators of $\mathfrak{gl}(n)$ on $L(\lambda)$ is given by the* Gelfand-Tsetlin formulas*:*

$$e_{k,k+1}(T(v)) = -\sum_{i=1}^{k} \left(\frac{\prod_{j=1}^{k+1}(v_{ki} - v_{k+1,j})}{\prod_{j\neq i}^{k}(v_{ki} - v_{kj})} \right) T(v + \delta^{ki}),$$

$$e_{k+1,k}(T(v)) = \sum_{i=1}^{k} \left(\frac{\prod_{j=1}^{k-1}(v_{ki} - v_{k-1,j})}{\prod_{j\neq i}^{k}(v_{ki} - v_{kj})} \right) T(v - \delta^{ki}),$$

$$e_{kk}(T(v)) = \left(k - 1 + \sum_{i=1}^{k} v_{ki} - \sum_{i=1}^{k-1} v_{k-1,i} \right) T(v).$$

If $e_{k,k+1}(T(v))$ or $e_{k+1,k}(T(v))$ contains a summand with a non-standard $T(v \pm \delta^{ki})$, then the summand is assumed to be zero.

These formulas define a Gelfand-Tsetlin modules where the action of the generators of Γ is given by:

$$c_{mk}(T(v)) = \gamma_{mk}(v) T(v),$$

where

$$\gamma_{mk}(v) := \sum_{i=1}^{m} (v_{mi} + m - 1)^k \prod_{j\neq i} \left(1 - \frac{1}{v_{mi} - v_{mj}} \right). \tag{2}$$

To every $w \in v + T_{n-1}(\mathbb{Z})$ we associate the maximal ideal of Λ generated by $\lambda_{ij} - w_{ij}$ and the maximal ideal $\mathbf{m}_w$ of Γ generated by $c_{ij} - \gamma_{ij}(w)$, where $\gamma_{ij}(w)$ are symmetric polynomials defined in (2). Again, the correspondence $w \mapsto \mathbf{m}_w$ is not one-to-one, a given $\mathbf{m} \in \operatorname{Specm}\Gamma$ defines the finite fiber of maximal ideals of Λ corresponding to the set $\widehat{\mathbf{m}}$ of $w \in v + T_{n-1}(\mathbb{Z})$ with $\mathbf{m}_w = \mathbf{m}$.

The basis of tableaux defined in Theorem 4.1 is called the *Gelfand-Tsetlin basis*. Discovery of Gelfand-Tsetlin bases are among the most remarkable results of the representation theory of classical Lie algebras. It provides a convenient realization of every simple finite dimensional representation of the Lie algebra $\mathfrak{gl}_n$. For other types of simple finite dimensional Lie algebras we refer to Molev [2006].

4.2 $U(\mathfrak{gl}_n)$ **is a Galois order over** Γ**.** We identify $T_{n-1}(\mathbb{Z})$ with the free abelian group $\mathcal{M} \simeq \mathbb{Z}^{\frac{n(n-1)}{2}}$ generated by δ^{ij}, $1 \leqslant j \leqslant i \leqslant n-1$, where $(\delta^{ij})_{ij} = 1$ and all other $(\delta^{ij})_{k\ell}$ are zero, $1 \leq j \leq i \leq n-1$. The group $\mathcal{M}$ acts naturally on $T_n(\mathbb{C})$ by translations. We also have the action of $G = S_n \times S_{n-1} \times \cdots \times S_1$ on $T_n(\mathbb{C})$ as follows:

$$\sigma(v) := (v_{n,\sigma^{-1}[n](1)}, \ldots, v_{n,\sigma^{-1}[n](n)} | \ldots | v_{1,\sigma^{-1}[1](1)}).$$

where $v \in T_n(\mathbb{C})$, $\sigma \in G$ and $\sigma[k] \in S_k$. This leads to the action of the semidirect product $G \ltimes T_{n-1}(\mathbb{Z})$ on $T_n(\mathbb{C})$.

Denote by K the field of fractions of Γ and by L the field of fractions of Λ. We have $L^G = K$, $\Lambda^G = \Gamma$ and $G = G(L/K)$ is the Galois group of the field extension $K \subset L$. The group $\mathcal{M}$ acts naturally on L and U is a subalgebra of $(L * \mathcal{M})^G$. Following Futorny and Ovsienko [2014], define a linear map $\tau : U \longmapsto (L * \mathcal{M})^G$ where

$$\tau(e_{mm}) = e_{mm} * e,\ \tau(e_{m\,m+1}) = \sum_{i=1}^{m} a_{mi}^{+} \delta^{mi},\ \tau(e_{m+1\,m}) = \sum_{i=1}^{m} a_{mi}^{-} (\delta^{mi})^{-1},$$

where

$$a_{mi}^{\pm} = \mp \frac{\prod_j (\lambda_{m\pm 1,j} - \lambda_{mi})}{\prod_{j \neq i} (\lambda_{mj} - \lambda_{mi})},$$

and e is the identity element of the group $\mathcal{M}$.

In fact, the map τ is algebra homomorphism since the defining relations of $\mathfrak{gl}_n$ are given by some rational functions which agree on finite dimensional modules, thus relations are satisfied.

Moreover, we have

Theorem 4.2 (Futorny and Ovsienko [2010], Proposition 7.2)**.** *τ is an embedding and U is a Galois order over Γ.*

Applying Theorem 3.2 we obtain that the number of isomorphism classes of simple Gelfand-Tsetlin U-modules with a given maximal ideal of Γ in the support is bounded by $\prod_{i=1}^{n-1} i!$. Another consequence of Theorem 4.2 is the following. If M is a Gelfand-Tsetlin U-module and $\mathbf{m} = \mathbf{m}_v \in \operatorname{Specm}\Gamma$ for some $v \in T_n(\mathbb{C})$ then

$$e_{k,k\pm 1} M(\mathbf{m}) \subset \sum_{i=1}^{k} M(\mathbf{m}_{v\pm\delta^{ki}}).$$

From here and Theorem 3.1 one easily obtains

Corollary 4.1. *(i)* Γ *is a Harish-Chandra subalgebra of* U.

(ii) Γ *is maximal commutative in* U.

4.3 Tableaux modules. The explicit nature of the Gelfand-Tsetlin formulas in Theorem 4.1 and the fact that the coefficients in the formulas are rational functions on the entries of the tableaux, naturally raises the question whether this construction can be extended for more general tableaux.

If V is a Gelfand-Tsetlin modules which has a basis parametrized by the tableaux and the action of Γ is determined by the entries of tableaux as in (2) then such V will be called *tableau module*. The problem of constructing of tableaux modules was studied by Gelfand and Graev in Gelfand and Graev [1965] and by Lemire and Patera (for $n = 3$) in Lemire and Patera [1979], Lemire and Patera [1985]. Tableux relalization for Generalzied Verma modules was considered in Mazorchuk [1998].

If the action of the generators of $\mathfrak{gl}_n$ on a tableau Gelfand-Tsetlin module V is given by the classical Gelfand-Tsetlin formulas as in Theorem 4.1 then V will be called *standard tableau module*. Modules considered in Gelfand and Graev [1965], Lemire and Patera [1979], Lemire and Patera [1985] are standard tableau modules.

We call $T(v)$ a *generic tableau* and v a *generic vector* if $v_{rs} - v_{rt} \notin \mathbb{Z}$ for any $r < n$ and all possible $s \neq t$. For a generic tableau all denominators in the Gelfand-Tsetlin formulas are nonintegers and one can use the same formulas to define *generic* standard tableau Gelfand-Tsetlin module $V(T(v))$ (Drozd, Futorny, and Ovsienko [1994], Section 2.3). All Gelfand-Tsetlin multiplicities of maximal ideals of $V(T(v))$ are 1.

Definition 4.1. *For each generic vector w and any $1 \leq r, s \leq n$ define*

$$d_{rs}(w) := \begin{cases} -\frac{\prod_{j=1}^{s}(w_{s-1,1}-w_{s,j})}{\prod_{j=2}^{s-1}(w_{s-1,1}-w_{s-1,j})} \prod_{j=r}^{s-2} \left(\frac{\prod_{t=2}^{j+1}(w_{j1} - w_{j+1,t})}{\prod_{t=2}^{j}(w_{j1} - w_{jt})} \right), & \text{if } r < s, \\ \frac{\prod_{j=1}^{s-1}(w_{s1}-w_{s-1,j})}{\prod_{j=2}^{s}(w_{s1}-w_{sj})} \prod_{j=s+2}^{r} \left(\frac{\prod_{t=2}^{j-2}(w_{j-1,1} - w_{j-2,t})}{\prod_{t=2}^{j-1}(w_{j-1,1} - w_{j-1,t})} \right), & \text{if } r > s, \\ r - 1 + \sum_{i=1}^{r} w_{ri} - \sum_{i=1}^{r-1} w_{r-1,i}, & \text{if } r = s, \end{cases}$$

Let $1 \leq r < s \leq n-1$. Set $\varepsilon_{rs} := \delta^{r,1} + \delta^{r+1,1} + \ldots + \delta^{s-1,1} \in T_n(\mathbb{Z})$, $\varepsilon_{rr} = 0$ and $\varepsilon_{sr} = -\varepsilon_{rs}$.

Let $\widetilde{S}_k$ be the subset of S_n consisting of the transpositions $(1, i)$, $i = 1, \ldots, k$. For $s < \ell$, set $\Phi_{s\ell} = \widetilde{S}_{\ell-1} \times \cdots \times \widetilde{S}_s$. For $s > \ell$ we set $\Phi_{s\ell} = \Phi_{\ell s}$. Finally we let $\Phi_{\ell\ell} = \{\mathrm{Id}\}$. Every σ in $\Phi_{s\ell}$ will be written as a $|s - \ell|$-tuple of transpositions $\sigma[k]$ (where $\sigma[k]$ is the k-th component of σ).

Proposition 4.1 (Futorny, Grantcharov, and Ramirez [2015]). *Let $v \in T_n(\mathbb{C})$ be generic. Then the $\mathfrak{gl}_n$-module structure on $V(T(v))$ is defined by the formulas:*

$$e_{k\ell}(T(v+z)) = \sum_{\sigma \in \Phi_{k\ell}} d_{k\ell}(\sigma(v+z)) T(v+z+\sigma(\varepsilon_{k\ell})), \tag{3}$$

for $z \in T_{n-1}(\mathbb{Z})$ and $1 \leq k, \ \ell \leq n$. Moreover, $V(T(v))$ is a Gelfand-Tsetlin module with action of Γ given by the formulas (2).

Note that if $v - v' \in T_{n-1}(\mathbb{Z})$ then $V(T(v))$ and $V(T(v'))$ are isomorphic as vectors spaces but not necessarily as the $\mathfrak{gl}_n$-modules. Simple generic Gelfand-Tsetlin modules were described in Futorny, Grantcharov, and Ramirez [ibid.].

The main difficulty in the defining of a tableau Gelfand-Tsetlin module structure on $V(T(v))$ is the existence of entries in one row of $T(v)$ that have integer difference. Let $v \in \mathbb{C}^{\frac{n(n+1)}{2}}$. A pair of entries (v_{ki_j}, v_{ki_s}) such that $k > 1$ and $v_{ki_j} - v_{ki_s} \in \mathbb{Z}$ is called a *singular pair*. We say that v (and $T(v)$) is *singuar* if v has singular pairs. First examples of infinite dimensional tableau Gelfand-Tsetlin modules with singular tableaux were considered in Gelfand and Graev [1965], Lemire and Patera [1979], Lemire and Patera [1985]. A new effective method of constructing standard tableau simple Gelfand - Tsetlin modules was proposed in Futorny, Ramirez, and Zhang [2016]. It allowed to obtain a large family of simple modules that have a basis consisting of Gelfand-Tsetlin tableaux and the action of the generators of $\mathfrak{gl}_n$ is given by the classical Gelfand-Tsetlin formulas. All examples obtained in Gelfand and Graev [1965], Lemire and Patera [1979], Lemire and Patera [1985] are particular cases of this construction. But the class of modules defined in

Futorny, Ramirez, and Zhang [2016] is more general. They build out of a tableau satisfying certain *FRZ-condition*.

A tableau $T(v)$ is *critical* if it has equal entries in one or more rows different from the top row. Otherwise, tableau is *noncritical*.

Theorem 4.3 (Futorny, Ramirez, and Zhang [ibid.], Theorem II)**.** *Let $T(w)$ be a tableau satisfying the FRZ-condition. There exists a unique simple Gelfand-Tsetlin $\mathfrak{gl}_n$-module V_w having the following properties:*

(i) $V_w(\mathbf{m}_w) \neq 0$*;*

(ii) V_m *has a basis consisting of noncritical tableaux and the action of the generators of* $\mathfrak{gl}_n$ *is given by the classical Gelfand-Tsetlin formulas (4.1).*

(iii) All Gelfand-Tsetlin multiplicities of maximal ideals of Γ *in the support of* V_w *equal* 1.

Theorem 4.3 provides a combinatorial way to explicitly construct a large class of infinite dimensional simple Gelfand-Tsetlin modules.

Conjecture 1. If V is a simple Gelfand-Tsetlin $\mathfrak{gl}_n$-module which has a basis consisting of noncritical tableaux with classical action of the generators of $\mathfrak{gl}_n$ then $V \simeq V_w$ for some w satisfying the FRZ-condition.

The conjecture holds for $n \leq 4$.

A systematic study of singular modules was initiated in Futorny, Grantcharov, and Ramirez [2015] where the case of a singular tableau $T(v)$ with a unique singular pair was considered (*1-singular* case). A significant difference from all previous cases is the existence of *derivative* tableaux in the basis of $V(T(v))$ which reflects the fact that the exact bound for the Gelfand-Tsetlin multiplicities of $V(T(v))$ is 2. Alternative interpretation of a tableau Gelfand-Tsetlin module structure on $V(T(v))$ in 1-singular case was given independently in Vishnyakova [2018] and Zadunaisky [2017]. Simple subquotients of $V(T(v))$ were described in Gomes and Ramirez [2016].

We say that v is *singular of index* $m \geq 2$ if:

(i) there exists a row k, $1 < k < n$, and m entries $v_{ki_1}, \ldots, v_{ki_m}$ on this row such that $v_{ki_j} - v_{ki_s} \in \mathbb{Z}$ for all $j, s \in \{1, \ldots, m\}$;

(ii) m is maximal with the property (i).

The case of arbitrary singularity of index $m = 2$ was solved in Futorny, Grantcharov, and Ramirez [2016a]. In this case any number of singular pairs (but not singular triples) and multiple singular pairs in the same row were allowed. Finally, the general case of an arbitrary singularity was solved in Vishnyakova [2017] (for p-singularity) and Ramírez and Zadunaisky [2017] (for arbitrary singularity). We provide the construction from Ramírez and Zadunaisky [ibid.] whose spirit is closer to our original approach.

Recall that L is the field of rational functions in λ_{ij}, $i = 1, \dots, n$, $j = 1, \dots, i$, and $a^{\pm}_{mi} \in L$ for all $1 \leq i \leq k < n$. Consider a set of all tableau with integral entries whose top row consists of zeros. We set $V_{\mathbb{C}}$ to be the $\mathbb{C}$-vector space with this basis, and $V_L = L \otimes_{\mathbb{C}} V_{\mathbb{C}}$. Since τ (4.2) is a homomorphism, V_L is a U-module, with the action of $\mathfrak{gl}_n$ given by the Gelfand-Tsetlin formulas.

The group G acts acts on V_L by the diagonal action, while $\mathcal{M}$ acts on Λ and L by translations: $\delta^{k,i} \cdot \lambda_{l,j} = \lambda_{l,j} + \delta_{k,l}\delta_{i,j}$.

Denote by A the algebra of regular functions over generic tableaux, that is those elements in L which can be evaluated in any generic tableau, and let V_A to be the A-submodule of V_L generated by all integral tableaux. Given a generic tableau $T(v)$, we can recover the corresponding generic module $V(T(v))$ by specializing V_A at v. If $T(v)$ is a singular tableau then we replace A with an algebra $B \subset L$ such that there exists a B-submodule $V_B \subset V_L$ which is also a U-submodule and any element of V_B can be evaluated at v. Specialization at v finally defines $V(T(v))$.

Each point $v \in \mathbb{C}^{\frac{n(n+1)}{2}}$ defines the following refinement $\eta(v)$ of v, which measures of how far is v from being generic Ramírez and Zadunaisky [ibid.]. Fix $1 \leq k \leq n-1$. Construct a graph with vertices $i = 1, \dots, k$, put an edge between i and j if and only if $v_{k,i} - v_{k,j}$ is integer. The graph is the disjoint union of connected components, we set $\eta^{(k)}$ to be a sequence of their cardinalities arranged in descending order. The entries $v_{k,i}$ that form one connected component are called an *η-block* of v. If v is generic then $\eta^{(k)} = (1^k)$, a sequence consisting of n ones. We set $\eta = (\eta^{(1)}, \dots, \eta^{(n-1)}, 1^n)$ to be the *η-type* of v and $\eta(v)$ to be the element in $\mathbb{C}^{\frac{n(n+1)}{2}}$ obtained from v by rearranging of it components to match the η-blocks.

Let B be the localization of Λ by the multiplicative subset of Λ generated by the elements

$$\lambda_{k,i} - \lambda_{k,j} - z,\ 1 \leq i < j \leq k < n, z \in \mathbb{Z} \setminus \{0\}.$$

Following Ramírez and Zadunaisky [ibid.], we say that v is in an *η-normal form* if $v_{k,i} - v_{k,j} \in \mathbb{Z}_{\geq 0}$ implies that $v_{k,i}$ and $v_{k,j}$ belong to the same η-block of v and $i < j$. Clearly, the orbit $G \cdot v$ has at least one (but not necessarily unique) element in normal form. We also say that v is an *η-critical* if it is in η-normal form and $v_{k,i} - v_{k,j} \in \mathbb{Z}$ implies $v_{k,i} = v_{k,j}$.

Consider a subgroup $G_\eta \subset G$ consisting of those elements of G which preserve the block structure of $\eta(v)$.

Fix η and an η-critical v. Set B_η to be the localization of B by the multiplicative set generated by all $\lambda_{k,i} - \lambda_{k,j}$ such that $v_{k,i} \neq v_{k,j}$.

Now we consider divided difference operators that play a key role in the construction of $V(T(v))$. Denote by $s_i^{(k)}$ the simple transposition in G_η which interchanges $\lambda_{k,i}$ and $\lambda_{k,i+1}$ and fixes all other elements. The *divided difference* associated to $s_i^{(k)}$ is

$$\partial_i^{(k)} = \frac{1}{\lambda_{k,i} - \lambda_{k,i+1}}(id - s_i^{(k)}).$$

It can be viewed as an element of the smash product $L\#G_\eta$, where $(f \otimes \sigma)\cdot(g \otimes \sigma') = f\sigma(g) \otimes \sigma\sigma'$ for $f, g \in L$ and all $\sigma, \sigma' \in G_\eta$.

Let $G_{\eta,k} \subset G_\eta$ be the corresponding component of S_k, $k = 1, \dots, n$. If $\sigma = s_{i_1}^{(k)} s_{i_2}^{(k)} \cdots s_{i_l}^{(k)}$ is a reduced decomposition for $\sigma \in G_{\eta,k}$ then set $\partial_\sigma = \partial_{i_1}^{(k)} \cdot \partial_{i_2}^{(k)} \cdots \partial_{i_l}^{(k)}$ which does not depend of the chosen reduced decomposition. This naturally extends to the whole group G_η.

For each $\sigma \in G_\eta$ we define the *symmetrized divided difference operator*

$$D_\sigma^\eta = \mathrm{sym}_\eta \cdot \partial_\sigma,$$

where $\mathrm{sym}_\eta = \frac{1}{|G_\eta|} \sum_{\sigma \in G_\eta} \sigma$.

Since B_η is closed under the action of G_η, we have $D_\sigma^\eta(f) \in B_\eta$ for all $f \in B_\eta$. Denote by $V_\eta \subset V_L$ the B_η-span of $\{D_\sigma^\eta T(z) \mid \sigma \in G_\eta, z \in T_{n-1}(\mathbb{Z})\}$. Then V_η is a U-submodule of V_L. Denote $\mathfrak{N}_\eta = \{z \in T_{n-1}(\mathbb{Z}) \mid v + z \text{ is in normal form}\}$. If $z \in \mathfrak{N}_\eta$ then the stabilizer subgroup of z in G_η is $G_{\epsilon(z)}$ where $\epsilon(z)$ is some refinement of η. Fix $z \in \mathfrak{N}$. We say that $\sigma \in G_\eta$ is a $\epsilon(z)$-shuffle if it is increasing in each $\epsilon(z)$-block. We denote the set of all $\epsilon(z)$-shuffles in G_η by $\mathrm{Shuffle}_{\epsilon(z)}^\eta$. Write $\bar{D}_\sigma^\eta(v+z) = 1 \otimes D_\sigma^\eta(z)$ for $z \in \mathfrak{N}$ and $\sigma \in \mathrm{Shuffle}_{\epsilon(z)}^\eta$.

Combining Theorems 5.3 and 5.6 of Ramírez and Zadunaisky [2017] we obtain

Theorem 4.4. *Let $V(T(v)) = \mathbb{C} \otimes_{B_\eta} V_\eta$, where $\mathbb{C}$ is a right B_η-module such that $1\dot{f} = f(v)$. Then $V(T(v))$ is a Gelfand-Tsetlin module with a basis $\{\bar{D}_\sigma^\eta(v+z) \mid z \in \mathfrak{N}_\eta, \sigma \in \mathrm{Shuffle}_{\epsilon(z)}^\eta\}$ and $\mathbf{m}_v$ belongs to the support of $V(T(v))$.*

Conjecture 2. Any simple Gelfand-Tsetlin module V with $V(\mathbf{m}_v) \neq 0$ is isomorphic to a subquotient of $V(T(v))$ for any singular v.

The conjecture was stated for any singular v of index 2 in Futorny, Grantcharov, and Ramirez [2016a]. It is known to be true for $n = 2$ and $n = 3$, and for the 1-singular v Futorny, Grantcharov, and Ramirez [2017]. In particular, it gives a complete classification of all simple Gelfand-Tsetlin $\mathfrak{gl}(3)$-modules, Futorny, Grantcharov, and Ramirez [2014].

References

V. V. Bavula (1992). "Generalized Weyl algebras and their representations". *Algebra i Analiz* 4.1, pp. 75–97. MR: 1171955 (cit. on pp. 1322, 1325, 1326).

V. V. Bavula and F. van Oystaeyen (2004). "Simple modules of the Witten-Woronowicz algebra". *J. Algebra* 271.2, pp. 827–845. MR: 2025552 (cit. on p. 1322).

Richard E. Block (1981). "The irreducible representations of the Lie algebra $\mathfrak{sl}(2)$ and of the Weyl algebra". *Adv. in Math.* 39.1, pp. 69–110. MR: 605353 (cit. on p. 1321).

Jonathan Brundan and Alexander Kleshchev (2006). "Shifted Yangians and finite W-algebras". *Adv. Math.* 200.1, pp. 136–195. MR: 2199632.

Mark Colarusso and Sam Evens (2010). "On algebraic integrability of Gelfand-Zeitlin fields". *Transform. Groups* 15.1, pp. 46–71. MR: 2600695 (cit. on p. 1321).

– (2014). "The Gelfand-Zeitlin integrable system and K-orbits on the flag variety". In: *Symmetry: representation theory and its applications*. Vol. 257. Progr. Math. Birkhäuser/Springer, New York, pp. 85–119. MR: 3363008 (cit. on p. 1321).

Jacques Dixmier (1974). *Algèbres enveloppantes*. Cahiers Scientifiques, Fasc. XXXVII. Gauthier-Villars Éditeur, Paris-Brussels-Montreal, Que., pp. ii+349. MR: 0498737 (cit. on p. 1322).

Yu. A. Drozd, Vyacheslav M. Futorny, and Serge A. Ovsienko (1994). "Harish-Chandra subalgebras and Gelfand-Zetlin modules". In: *Finite-dimensional algebras and related topics (Ottawa, ON, 1992)*. Vol. 424. NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. Kluwer Acad. Publ., Dordrecht, pp. 79–93. MR: 1308982 (cit. on pp. 1321–1323, 1326, 1332).

Yu. A. Drozd, Serge A. Ovsienko, and Vyacheslav M. Futorny (1989). "Irreducible weighted sl(3)-modules". *Funkts. Anal. i Prilozhen.* 23.3, pp. 57–58. MR: 1026989 (cit. on p. 1321).

– (1991). "On Gelfand-Zetlin modules". In: *Proceedings of the Winter School on Geometry and Physics (Srní, 1990)*. 26, pp. 143–147. MR: 1151899 (cit. on pp. 1321, 1322, 1326, 1328).

A. G. Elashvili and V. G. Kac (2005). "Classification of good gradings of simple Lie algebras". In: *Lie groups and invariant theory*. Vol. 213. Amer. Math. Soc. Transl. Ser. 2. Amer. Math. Soc., Providence, RI, pp. 85–104. MR: 2140715.

S. L. Fernando (1990). "Lie algebra modules with finite-dimensional weight spaces. I". *Trans. Amer. Math. Soc.* 322.2, pp. 757–781. MR: 1013330 (cit. on p. 1321).

Vyacheslav M. Futorny, Dimitar Grantcharov, and Luis Enrique Ramirez (2014). "On the classification of irreducible Gelfand-Tsetlin modules of $\mathfrak{sl}(3)$". In: *Recent advances in representation theory, quantum groups, algebraic geometry, and related topics*. Vol. 623. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 63–79. MR: 3288623 (cit. on pp. 1322, 1336).

Vyacheslav M. Futorny, Dimitar Grantcharov, and Luis Enrique Ramirez (2015). "Irreducible generic Gelfand-Tsetlin modules of $\mathfrak{gl}(n)$". *SIGMA Symmetry Integrability Geom. Methods Appl.* 11, Paper 018, 13. MR: 3322336 (cit. on pp. 1322, 1333, 1334).

– (June 2016a). "New singular Gelfand-Tsetlin $\mathfrak{gl}(n)$-modules of index 2". To appear in Communications in Mathematical Physics. arXiv: 1606.03394 (cit. on pp. 1322, 1334, 1336).

– (2016b). "Singular Gelfand-Tsetlin modules of $\mathfrak{gl}(n)$". *Adv. Math.* 290, pp. 453–482. MR: 3451929 (cit. on p. 1322).

– (Apr. 2017). "Drinfeld category and the classification of singular Gelfand-Tsetlin $\mathfrak{gl}_n$-modules". To appear in IMRN. arXiv: 1704.01209 (cit. on p. 1336).

Vyacheslav M. Futorny, Alexander I. Molev, and Serge A. Ovsienko (2005). "Harish-Chandra modules for Yangians". *Represent. Theory* 9, pp. 426–454. MR: 2142818 (cit. on pp. 1322, 1324, 1326, 1328).

Vyacheslav M. Futorny and Serge A. Ovsienko (2005). "Kostant's theorem for special filtered algebras". *Bull. London Math. Soc.* 37.2, pp. 187–199. MR: 2119018 (cit. on p. 1324).

– (2007). "Harish-Chandra categories". In: *Proceedings of the XVIth Latin American Algebra Colloquium (Spanish)*. Bibl. Rev. Mat. Iberoamericana. Rev. Mat. Iberoamericana, Madrid, pp. 271–286. MR: 2500363 (cit. on p. 1321).

– (2010). "Galois orders in skew monoid rings". *J. Algebra* 324.4, pp. 598–630. MR: 2651560 (cit. on pp. 1322, 1325–1327, 1331).

– (2014). "Fibers of characters in Gelfand-Tsetlin categories". *Trans. Amer. Math. Soc.* 366.8, pp. 4173–4208. MR: 3206456 (cit. on pp. 1322, 1327, 1331).

Vyacheslav M. Futorny, Luis Enrique Ramirez, and Jian Zhang (Nov. 2016). "Combinatorial construction of Gelfand-Tsetlin modules for $\mathfrak{gl}_n$". arXiv: 1611.07908 (cit. on pp. 1322, 1333, 1334).

Pierre Gabriel (1962). "Des catégories abéliennes". *Bull. Soc. Math. France* 90, pp. 323–448. MR: 0232821.

I. M. Gelfand and M. L. Cetlin (1950). "Finite-dimensional representations of groups of orthogonal matrices". *Doklady Akad. Nauk SSSR (N.S.)* 71, pp. 1017–1020. MR: 0034763 (cit. on pp. 1321, 1329, 1330).

I. M. Gelfand and M. I. Graev (1965). "Finite-dimensional irreducible representations of the unitary and complete linear group and special functions associated with them". *Izv. Akad. Nauk SSSR Ser. Mat.* 29, pp. 1329–1356. MR: 0201568 (cit. on pp. 1332, 1333).

Carlos Alexandre Gomes and Luis Enrique Ramirez (Dec. 2016). "Families of irreducible singular Gelfand-Tsetlin modules of $\mathfrak{gl}(n)$". arXiv: 1612.00636 (cit. on p. 1334).

M. I. Graev (2004). "Infinite-dimensional representations of the Lie algebra $\mathfrak{gl}(n, \mathbb{C})$ related to complex analogs of the Gelfand-Tsetlin patterns and general hypergeometric functions on the Lie group $\mathrm{GL}(n, \mathbb{C})$". *Acta Appl. Math.* 81.1-3, pp. 93–120. MR: 2069333 (cit. on p. 1321).

– (2007). "A continuous analogue of Gelfand-Tsetlin schemes and a realization of the principal series of irreducible unitary representations of the group $\mathrm{GL}(n, \mathbb{C})$ in the space of functions on the manifold of these schemes". *Dokl. Akad. Nauk* 412.2, pp. 154–158. MR: 2451955 (cit. on p. 1321).

V. Guillemin and S. Sternberg (1983). "The Gelfand-Cetlin system and quantization of the complex flag manifolds". *J. Funct. Anal.* 52.1, pp. 106–128. MR: 705993 (cit. on p. 1321).

Jonas T. Hartwig (Oct. 2017a). "Principal Galois orders and Gelfand-Zeitlin modules". arXiv: 1710.04186 (cit. on pp. 1322, 1327, 1328).

– (2017b). "The q-difference Noether problem for complex reflection groups and quantum OGZ algebras". *Comm. Algebra* 45.3, pp. 1166–1176. MR: 3573367 (cit. on p. 1328).

Bertram Kostant and Nolan Wallach (2006a). "Gelfand-Zeitlin theory from the perspective of classical mechanics. I". In: *Studies in Lie theory*. Vol. 243. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 319–364. MR: 2214253 (cit. on p. 1321).

– (2006b). "Gelfand-Zeitlin theory from the perspective of classical mechanics. II". In: *The unity of mathematics*. Vol. 244. Progr. Math. Birkhäuser Boston, Boston, MA, pp. 387–420. MR: 2181811 (cit. on p. 1321).

F. Lemire and J. Patera (1979). "Formal analytic continuation of Gelfand's finite-dimensional representations of $\mathrm{gl}(n, \mathbf{C})$". *J. Math. Phys.* 20.5, pp. 820–829. MR: 531284 (cit. on pp. 1332, 1333).

– (1985). "Gelfand representations of $\mathrm{sl}(n, \mathbf{C})$". *Algebras Groups Geom.* 2.2, pp. 146–166. MR: 901579 (cit. on pp. 1332, 1333).

Olivier Mathieu (2000). "Classification of irreducible weight modules". *Ann. Inst. Fourier (Grenoble)* 50.2, pp. 537–592. MR: 1775361 (cit. on p. 1321).

Volodymyr Mazorchuk (1998). "Tableaux realization of generalized Verma modules". *Canad. J. Math.* 50.4, pp. 816–828. MR: 1638623 (cit. on pp. 1321, 1332).

– (1999). "Orthogonal Gelfand-Zetlin algebras. I". *Beiträge Algebra Geom.* 40.2, pp. 399–415. MR: 1720114 (cit. on p. 1328).

– (2001). "On categories of Gelfand-Zetlin modules". In: *Noncommutative structures in mathematics and physics (Kiev, 2000)*. Vol. 22. NATO Sci. Ser. II Math. Phys. Chem. Kluwer Acad. Publ., Dordrecht, pp. 299–307. MR: 1893473 (cit. on p. 1321).

Volodymyr Mazorchuk and Lyudmyla Turowska (2000). "On Gelfand-Zetlin modules over $U_q(\mathrm{gl}_n)$". *Czechoslovak J. Phys.* 50.1. Quantum groups and integrable systems (Prague, 1999), pp. 139–144. MR: 1798777 (cit. on p. 1328).

J. C. McConnell and J. C. Robson (1987). *Noncommutative Noetherian rings*. Pure and Applied Mathematics (New York). With the cooperation of L. W. Small, A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, pp. xvi+596. MR: 934572 (cit. on p. 1325).

A. S. Miščenko and A. T. Fomenko (1978). "Euler equation on finite-dimensional Lie groups". *Izv. Akad. Nauk SSSR Ser. Mat.* 42.2, pp. 396–415, 471. MR: 0482832.

Alexander I. Molev (2006). "Gelfand-Tsetlin bases for classical Lie algebras". In: *Handbook of algebra. Vol. 4*. Vol. 4. Handb. Algebr. Elsevier/North-Holland, Amsterdam, pp. 109–170. MR: 2523420 (cit. on pp. 1321, 1331).

Andrei Okounkov and Anatoly Vershik (1996). "A new approach to representation theory of symmetric groups". *Selecta Math. (N.S.)* 2.4, pp. 581–605. MR: 1443185 (cit. on p. 1324).

Serge A. Ovsienko (2002). "Finiteness statements for Gelfand–Tsetlin modules". In: Algebraic structures and their applications. Kiev: Math. Inst. (cit. on p. 1322).

Luis Enrique Ramírez and Pablo Zadunaisky (May 2017). "Gelfand-Tsetlin modules over $\mathfrak{gl}(n, \mathbb{C})$ with arbitrary characters". arXiv: 1705.10731 (cit. on pp. 1322, 1335, 1336).

Elizaveta Vishnyakova (May 2017). "Geometric approach to p-singular Gelfand-Tsetlin $\mathfrak{gl}_n$-modules". arXiv: 1705.05793 (cit. on pp. 1322, 1335).

– (2018). "A geometric approach to 1-singular Gelfand–Tsetlin $\mathfrak{gl}_n$-modules". *Differential Geom. Appl.* 56, pp. 155–160. arXiv: 1704.00170. MR: 3759359 (cit. on pp. 1322, 1334).

Pablo Zadunaisky (2017). "A new way to construct 1-singular Gelfand-Tsetlin modules". *Algebra Discrete Math.* 23.1, pp. 180–193. MR: 3634504 (cit. on pp. 1322, 1334).

D. P. Želobenko (1973). *Compact Lie groups and their representations*. Translated from the Russian by Israel Program for Scientific Translations, Translations of Mathematical Monographs, Vol. 40. American Mathematical Society, Providence, R.I., pp. viii+448. MR: 0473098 (cit. on pp. 1321, 1328).

Received 2017-12-06.

Vyacheslav Futorny
Instituto de Matemática e Estatística
Universidade de São Paulo
Caixa Postal 66281
05315-970 São Paulo
Brasil
futorny@ime.usp.br

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1339–1362)

A VIEW ON INVARIANT RANDOM SUBGROUPS AND LATTICES

Tsachik Gelander

Abstract

For more than half a century lattices in Lie groups played an important role in geometry, number theory and group theory. Recently the notion of Invariant Random Subgroups (IRS) emerged as a natural generalization of lattices. It is thus intriguing to extend results from the theory of lattices to the context of IRS, and to study lattices by analyzing the compact space of all IRS of a given group. This article focuses on the interplay between lattices and IRS, mainly in the classical case of semisimple analytic groups over local fields.

Let G be a locally compact group. We denote by $\mathrm{Sub}(G)$ the space of closed subgroups of G equipped with the Chabauty topology. The compact space $\mathrm{Sub}(G)$ is usually too complicated to work with directly. However, considering a random point in $\mathrm{Sub}(G)$ is often much more effective. Note that G acts on $\mathrm{Sub}(G)$ by conjugation. An invariant random subgroup (or shortly IRS) is a G-invariant probability measure on $\mathrm{Sub}(G)$. We denote by $\mathrm{IRS}(G)$ the space of all IRSs of G equipped with the w^*-topology. By Riesz' representation theorem and Alaoglu's theorem, $\mathrm{IRS}(G)$ is compact.

The Dirac measures in $\mathrm{IRS}(G)$ correspond to normal subgroups. Any lattice Γ in G induces an IRS μ_Γ which is defined as the push forward of the G-invariant probability measure from G/Γ to $\mathrm{Sub}(G)$ via the map $g\Gamma \mapsto g\Gamma g^{-1}$.

More generally consider a probability measure preserving action $G \curvearrowright (X, m)$. By a result of Varadarajan, the stabilizer of almost every point in X is closed in G. Moreover, the stabilizer map $X \to \mathrm{Sub}_G,\ x \mapsto G_x$ is measurable, and hence one can push the measure m to an IRS on G. In other words the random subgroup is the stabilizer of a random point in X. In a sense, the study of pmp G-spaces can be divided to the study of stabilizers (i.e. IRSs), the study of orbit spaces and the interplay between the two. Vice versa, every IRS arises (non-uniquely) in this way (see Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a, Theorem 2.6]).

MSC2010: primary 22D05; secondary 22E40, 22B99, 20G15, 37A05, 53C35.

Since its rebirth in the beginning of the current decade (see Section 10 for a short summery of the history of IRS), the topic of IRS played an important role in various parts of group theory, geometry and dynamics, and attracted the attention of many mathematicians for various different reasons. Not aiming to give an overview, I will try to highlight here several aspects of the evolving theory.

1 IRS and Lattices

IRSs can be considered as a generalisation of lattices, and one is tempted to extend results from the theory of lattices to IRS. In the other direction, as often happens in mathematics where one considers random objects to prove result about deterministic ones, the notion of IRS turns out to yield an extremely powerful tool to study lattices. In this section I will try to give a taste of the interplay between IRSs and Lattices focusing mainly on the second point of view. Attempting to expose the phenomenon in a rather clear way, avoiding technicality, I will assume throughout most of this section that G is a noncompact simple Lie group, although most of the results can be formulated in the much wider setup of semisimple analytic groups over arbitrary local fields, see Section 1.5.

1.1 Borel Density. Let $\mathrm{PSub}(G)$ denote the space of proper closed subgroups. Since G is an isolated point in $\mathrm{Sub}(G)$ (see Toyama [1949] and Kuranishi [1951]) we deduce that $\mathrm{PSub}(G)$ is compact. Letting $\mathrm{PIRS}(G)$ denote the subspace of $\mathrm{IRS}(G)$ consisting of the measures supported on $\mathrm{PSub}(G)$, we deduce:

Lemma 1.1. *The space of proper IRSs, $PIRS(G)$ is compact.*

Let us say that that an IRS μ is discrete if a random subgroup is μ almost surely discrete, and denote by $\mathrm{DIRS}(G)$ the subspace of $\mathrm{IRS}(G)$ consisting of discrete IRSs. The following is a generalization of the classical Borel Density Theorem:

Theorem 1.2. *(Borel Density Theorem for IRS, Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a, Theorem 2.9]) Every proper IRS in G is discrete, i.e. $PIRS(G) = DIRS(G)$. Moreover, for every $\mu \in DIRS(G)$, μ-almost every subgroup is either trivial or Zariski dense.*

In order to prove Theorem 1.2 one first observes that there are only countably many conjugacy classes of non-trivial finite subgroups in G, hence the measure of their union is zero with respect to any non-atomic IRS. Then one can apply the same idea as in Furstenberg's proof of the classical Borel density theorem Furstenberg [1976]. Indeed, taking the Lie algebra of $H \in \mathrm{Sub}(G)$ as well as of its Zariski closure induce measurable maps (see

Gelander and Levit [2017, §4])

$$H \mapsto \mathrm{Lie}(H),\ H \mapsto \mathrm{Lie}(\overline{H}^Z).$$

As G is noncompact, Furstenberg's argument implies that the Grassman variety of non-trivial subspaces of $\mathrm{Lie}(G)$ does not carry an $\mathrm{Ad}(G)$-invariant measure. It follows that $\mathrm{Lie}(H) = 0$ and $\mathrm{Lie}(\overline{H}^Z) \in \{\mathrm{Lie}(G), 0\}$ almost surely, and the two statements of the theorem follow.

1.2 Weak Uniform Discreteness.

Let U be an identity neighbourhood in G. A family of subgroups $\mathfrak{F} \subset \mathrm{Sub}(G)$ is called *U-uniformly discrete* if $\Gamma \cap U = \{1\}$ for all $\Gamma \in \mathfrak{F}$.

Definition 1.3. A family $\mathfrak{F} \subset \mathrm{DIRS}(G)$ of invariant random subgroups is said to be *weakly uniformly discrete* if for every $\epsilon > 0$ there is an identity neighbourhood $U_\epsilon \subset G$ such that

$$\mu(\{\Gamma \in \mathrm{Sub}_G : \Gamma \cap U_\epsilon \neq \{1\}\}) < \epsilon$$

for every $\mu \in \mathfrak{F}$.

A justification for this definition is given by the following result which is proved by an elementary argument and yet provides a valuable information:

Theorem 1.4. *Let G be a connected non-compact simple Lie group. Then DIRS(G) is weakly uniformly discrete.*

Let $U_n, n \in \mathbb{N}$ be a descending sequence of compact sets in G which form a base of identity neighbourhoods, and set

$$K_n = \{\Gamma \in \mathrm{Sub}_G : \Gamma \cap U_n = \{1\}\}.$$

Since G has NSS (no small subgroups), i.e. there is an identity neighbourhood which contains no non-trivial subgroups, we have:

Lemma 1.5. *The sets K_n are open in* $\mathrm{Sub}(G)$.

Proof. Fix n and let $V \subset U_n$ be an open identity neighbourhood which contains no non-trivial subgroups, such that $V^2 \subset U_n$. It follows that a subgroup Γ intersects U_n non-trivially iff it intersects $U_n \setminus V$. Since $U_n \setminus V$ is compact, the lemma is proved. □

In addition, observe that the ascending union $\bigcup_n K_n$ exhausts $\mathrm{Sub}_d(G)$, the set of all discrete subgroups of G. Therefore we have:

Claim 1.6. For every $\mu \in \mathrm{DIRS}(G)$ and $\epsilon > 0$ we have $\mu(K_n) > 1 - \epsilon$ for some n. □

Let

$$\mathcal{K}_{n,\epsilon} := \{\mu \in \mathrm{DIRS}(G) : \mu(K_n) > 1 - \epsilon\}.$$

Since $\mathrm{Sub}(G)$ is metrizable, it follows from Lemma 1.5 that $\mathcal{K}_{n,\epsilon}$ is open. By Claim 1.6, for any given $\epsilon > 0$, the sets $\mathcal{K}_{n,\epsilon}$, $n \in \mathbb{N}$ form an ascending cover of $\mathrm{DIRS}(G)$. Since the latter is compact, we have $\mathrm{DIRS}(G) \subset \mathcal{K}_{m,\epsilon}$ for some $m = m(\epsilon)$. It follows that

$$\mu\big(\{\Gamma \in \mathrm{Sub}(G) : \Gamma \text{ intersects } U_m \text{ trivially}\}\big) > 1 - \epsilon,$$

for every $\mu \in \mathrm{DIRS}(G)$. Thus Theorem 1.4 is proved. □

Picking $\epsilon < 1$ and applying the theorem for the IRS μ_Γ where $\Gamma \le G$ is an arbitrary lattice, one deduces the Kazhdan–Margulis theorem Každan and Margulis [1968], and in particular that there is a positive lower bound on the volume of locally G/K-orbifolds:

Corollary 1.7 (Kazhdan–Margulis theorem)**.** *There is an identity neighbourhood $\Omega \subset G$ such that for every lattice $\Gamma \le G$ there is $g \in G$ such that $g\Gamma g^{-1} \cap \Omega = \{1\}$.*

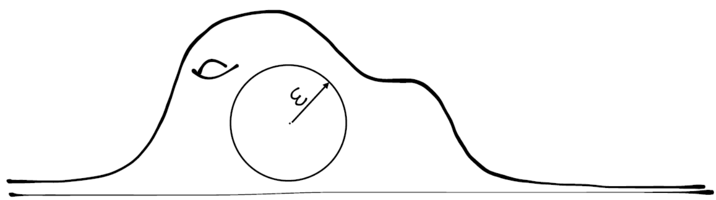


Figure 1: Every X-manifold has a thick part.

A famous conjecture of Margulis Margulis [1991, page 322] asserts that the set of all torsion-free anisotropic arithmetic lattices in G is U-uniformly discrete for some identity neighbourhood $U \subset G$. Theorem 1.4 can be regarded as a probabilistic variant of this conjecture as it implies that all lattices in G are jointly weakly uniformly discrete.

In the language of pmp actions Theorem 1.4 can be reformulated as follows:

Theorem 1.8 (p.m.p. actions are uniformly weakly locally free)**.** *For every $\epsilon > 0$ there is an identity neighbourhood $U_\epsilon \subset G$ such that the stabilizers of $1 - \epsilon$ of the points, in any non-atomic probability measure preserving G-space (X, m) are U_ϵ-uniformly discrete. I.e., there is a subset $Y \subset X$ with $m(Y) > 1 - \epsilon$ such that $U_\epsilon \cap G_y = \{1\}$, $\forall y \in Y$.*

1.3 Local Rigidity.

Observe that local rigidity implies Chabauty locally rigid:

Proposition 1.9. *Let G be a connected Lie group and $\Gamma \le G$ a locally rigid lattice. Then Γ is Chabauty locally rigid, i.e. the conjugacy class of Γ is Chabauty open.*

Proof. Let $\Gamma \le G$ a locally rigid lattice. Let U be a compact identity neighborhood in G satisfying:

- $U \cap \Gamma = \{1\}$,
- U contains no nontrivial groups,

and let V be an open symmetric identity neighborhood with $V^2 \subset U$. By the choice of V we for a subgroup $H \le G$, that $H \cap U \ne \{1\}$ iff H meets the compact set $U \setminus V$.

Recall that Γ, being a lattice in a Lie group, is finitely presented and let $\langle \Sigma | R \rangle$ be a finite presentation of Γ. Denote $S = \{s_1, \dots, s_k\}$. We can pick a sufficiently small identity neighborhood Ω so that for every choice of $g_i \in s_i\Omega$, $i = 1, \dots, k$ and every $w \in R$ we have $w(g_1, \dots, g_n) \in U$.

Now if $H \in \mathrm{Sub}(G)$ is sufficiently close to Γ in the Chabauty topology then $H \cap s_i\Omega \ne \emptyset$, $i = 1, \dots, k$ and $H \cap (U \setminus V) = \emptyset$, i.e. $H \cap U = \{1\}$. Picking $h_i \in H \cap s_i\Omega$, $i = 1, \dots, k$ one sees that the assignment $s_i \mapsto h_i$ induces a homomorphism from Γ into H. Since Γ is locally rigid it follows that if H is sufficiently close to Γ then it contains a conjugate of Γ. However there are only finitely many subgroups containing Γ and intersecting U trivially, hence if H is sufficiently close to Γ then it is a conjugate of Γ. □

Denote by $\mathrm{EIRS}(G)$ the space ergodic IRSs of G, i.e. the set of extreme points of $\mathrm{IRS}(G)$.

Corollary 1.10. *Let G be a connected Lie group and $\Gamma \le G$ a locally rigid lattice. Then the IRS μ_Λ is isolated in EIRS(G).*

Proof. Let Γ be as above. If μ is an IRS of G sufficiently close to μ_Γ then with positive μ-probability a random subgroup is a conjugate of Γ. Thus if μ is ergodic then it must be μ_Γ. □

1.4 Farber property.

Definition 1.11. A sequence μ_n of invariant random subgroups of G is called *Farber*[1] if μ_n converge to the trivial IRS, $\delta_{\{1\}}$.

[1] Various authors use various variants of this notion.

One of the key results of Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a] is the following theorem:

Theorem 1.12. *Let G be a simple Lie group of real rank at least 2. Let Γ_n be a sequence of pairwise non-conjugate lattices, then μ_{Γ_n} is Farber.*

The proof relies on the following variant of Stuck–Zimmer theorem (see Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [ibid.]):

Theorem 1.13. *Assuming $\mathrm{rank}(G) \geq 2$, a proper non-trivial ergodic IRS of G is μ_Γ for some lattice Γ, i.e.*

$$EIRS(G) = \{\delta_G, \delta_{\{1\}}, \mu_\Gamma,\ \Gamma \text{ a lattice in } G\}.$$

Proof of Theorem 1.12. Since G is of rank ≥ 2 it has Kazhdan's property (T), Každan [1967]. By E. Glasner and Weiss [1997], $\mathrm{EIRS}(G)$ is compact. Since δ_G and δ_{μ_Γ} where Γ is a lattice in G are isolated points, it follows that $\delta_{\{1\}}$ is the unique accumulation point of $\mathrm{EIRS}(G)$. In particular $\mu_{\Gamma_n} \to \delta_{\{1\}}$. □

1.5 Semisimple analytic groups. Above we have carried out the arguments under the assumption that G is a simple Lie group. In this section we state the results in the more general setup of analytic groups over local fields.

Definition 1. Let k be a local field and $\mathbb{G}$ a connected k-isotropic k-simple linear k-algebraic group.

- A *simple analytic group* is a group of the form $\mathbb{G}(k)$.
- A *semisimple analytic group* is an almost direct product of finitely many simple analytic groups, possibly over different local fields.

Note that if k is a local field and $\mathbb{G}$ is a connected semisimple linear k-algebraic group without k-anisotropic factors then $\mathbb{G}(k)$ is a semisimple analytic group. Such a group is indeed *analytic* in the sense of e.g. Serre [2006]. Associated to a semisimple analytic group G are its universal covering group $\widetilde{G}$ and adjoint group $\overline{G}$. There are central k-isogenies $\widetilde{G} \overset{\widetilde{p}}{\to} G \overset{\overline{p}}{\to} \overline{G}$ and this data is unique up to a k-isomorphism Margulis [1991, p. I.4.11]. For a semisimple analytic group G, denote by G^+ the subgroup of G generated by its unipotent elements Margulis [ibid., pp. I.1.5, I.2.3]. If G is simply connected then $G = G^+$. If G is Archimedean then G^+ is the connected component G_0 at the identity. In general G/G^+ is a compact abelian group. The group G^+ admits no proper finite index subgroups.

Definition 2. A simple analytic group G is *happy* if $\mathrm{char}(k)$ does not the divide $|Z|$ where Z is the kernel of the map $\widetilde{G} \to G$. A semisimple analytic group is *happy* if all of its almost direct factors are.

Note that a simply connected or a zero characteristic semisimple analytic group is automatically happy. From the work of Barnea and Larsen [2004] one obtains that a semisimple analytic group G is happy, iff G/G^+ is a finite abelian group, iff some (equivalently every) compact open subgroup in the non-Archimedean factor of G is finitely generated.

Self Chabauty isolation.

Definition 3. A l.c.s.c. group G is *self-Chabauty-isolated if the point G is isolated in* $\mathrm{Sub}(G)$ *with the Chabauty topology.*

Note that G is self-Chabauty-isolated if and only if there is a finite collection of open subsets $U_1, \ldots, U_n \subset G$ so that the only closed subgroup intersecting every U_i non-trivially is G itself. The following result is proved in Gelander and Levit [2017, §6].

Theorem 4. Let G be a happy semisimple analytic group. Then G^+ is self-Chabauty-isolated.

As an immediate consequence we deduce the analog of Lemma 1.1, namely that the space $\mathrm{PSub}(G^+)$ is compact for every G as in Theorem 4.

Borel Density. The following generalization of Theorem 1.2 was obtained in Gelander and Levit [ibid., Theorem 1.9]:

Theorem 5 (Borel density theorem for IRS)**.** Let k be a local field and G a happy semisimple analytic group over k. Assume that G has no almost k-simple factors of type B_n, C_n or F_4 if $\mathrm{char}(k) = 2$ and of type G_2 if $\mathrm{char}(k) = 3$.

Let μ be an ergodic invariant random subgroup of G. Then there is a pair of normal subgroups $N, M \lhd G$ so that

$$N \leq H \leq M, \quad H/N \text{ is discrete in } G/N \quad \text{and} \quad \overline{H}^{Z} = M$$

for μ-almost every closed subgroup H in G. Here $\overline{H}^{Z}$ is the Zariski closure of H.

Weak Uniform Discreteness. As shown in Gelander [2018, Theorem 2.1] the analog of Theorem 1.4 holds for general semisimple Lie groups:

Theorem 1.14. *Let G be a connected center-free semisimple Lie group with no compact factors. Then $\mathrm{DIRS}(G)$ is weakly uniformly discrete.*

Consider now a general locally compact σ-compact group G. Since $\mathrm{Sub}_d(G) \subset \mathrm{Sub}(G)$ is a measurable subset, by restricting attention to it, one may replace Property NSS by the weaker Property NDSS (no discrete small subgroups), which means that there is an identity neighbourhood which contains no non-trivial discrete subgroups. In that generality, the analog of Lemma 1.5 would say that K_n are relatively open in $\mathrm{Sub}_d(G)$. Thus, the ingredients required for the argument above are:

1. $\mathrm{DIRS}(G)$ is compact,

2. G has NDSS.

In particular we have:

Theorem 1.15. *Let G be a locally compact σ-compact group which satisfies (1) and (2). Then DIRS(G) is weakly uniformly discrete.*

If G possesses the Borel density theorem and G is self-Chabauty-isolated then (1) holds. By the previous paragraphs happy semisimple analytic groups enjoy these two properties, and hence (1).

***p*-adic groups.** Note that a p-adic analytic group G has NDSS, and hence $\mathrm{DIRS}(G)$ is uniformly discrete (in the obvious sense). Moreover, if $G \leq \mathrm{GL}_n(\mathbb{Q}_p)$ is a rational algebraic subgroup, then the first principal congruence subgroup $G(p\mathbb{Z}_p)$ is a torsion-free open compact subgroup. In particular the space $\mathrm{DIRS}(G)$ is $G(p\mathbb{Z}_p)$-uniformly discrete. Supposing further that G is simple, then in view of the Borel density theorem we have:

Let (X, μ) be a probability G-space essentially with no global fixed points. Then the action of the congruence subgroup $G(p\mathbb{Z}_p)$ on X is essentially free.

Positive characteristic. Algebraic groups over local fields of positive characteristic do not posses property NDSS, and the above argument does not apply to them.

Conjecture 1.16. *Let k be a local field of positive characteristic, let $\mathbb{G}$ be simply connected absolutely almost simple k-group with positive k-rank and let $G = \mathbb{G}(k)$ be the group of k-rational points. Then DIRS(G) weakly uniformly discrete.*

The analog of Corollary 1.7 in positive characteristic was proved in Salehi Golsefidy [2013] and Raghunathan [1972]. A. Levit proved Conjecture 1.16 for k-rank one groups Levit [n.d.].

Local Rigidity. Combining Theorem 7.2 and Proposition 7.9 of Gelander and Levit [2017] we obtain the following extension of Proposition 1.9:

Theorem 6. (Chabauty local rigidity Gelander and Levit [ibid.]) Let G be a semisimple analytic group and Γ an irreducible lattice in G. If Γ is locally rigid then it is also Chabauty locally rigid.

Let us also mention the following generalization of the classical Weil local rigidity theorem:

Theorem 7. (Gelander and Levit [2016, Theorem 1.2] Let X be a proper geodesically complete CAT(0) space without Euclidean factors and with Isom$X$ acting cocompactly. Let Γ be a uniform lattice in Isom$X$. Assume that for every de Rham factor Y of X isometric to the hyperbolic plane the projection of Γ to Isom$Y$ is non-discrete. Then Γ is locally rigid.

Farber Property. The proof presented above for Theorem 1.12 follows the lines developed at Gelander and Levit [2017] and is simpler and applies to a more general setup than the original proof from Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a]. In particular, the following general version of Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [ibid., Theorem 4.4] is proved in Gelander and Levit [2017]:

Theorem 8. Gelander and Levit [ibid., Theorem 1.1] Let G be a semisimple analytic group. Assume that G is happy, has property (T) and rank$(G) \geq 2$. Let Γ_n be a sequence of pairwise non-conjugate irreducible lattices in G. Then Γ_n is Farber.

Farber property for congruence subgroups. Relying on intricate estimates involving the trace formula, Raimbault [2013] and Fraczyk [2016] were able to establish Benjamini–Schramm convergence for every sequence of congruence lattices in the rank one groups $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{C})$.

Using property (τ) as a replacement for property (T), Levit [2017a] established Benjamini–Schramm convergence for every sequence of congruence lattices in any higher-rank semisimple group G over local fields. Whenever lattices in G are known to satisfy the congruence subgroup property this applies to all irreducible lattices in G.

1.6 The IRS compactification of moduli spaces. One may also use IRS(G) in order to obtain new compactifications of certain natural spaces.

Example 1.17. Let Σ be a closed surface of genus ≥ 2. Every hyperbolic structure on Σ corresponds to an IRS in $\mathrm{PSL}_2(\mathbb{R})$. Indeed one take a random point and a random tangent vector w.r.t the normalized Riemannian measure on the unit tangent bundle and consider the associated embedding of the fundamental $\pi_1(\Sigma)$ in $\mathrm{PSL}_2(\mathbb{R})$ via deck transformations.

Taking the closure in $\mathrm{IRS}(G)$ of the set of hyperbolic structures on Γ, one obtains an interesting compactification of the moduli space of Σ.

Problem 1.18. Analyse the IRS compactification of $\mathrm{Mod}(\Sigma)$.

Note that the resulting compactification is similar to (but is not exactly) the Deligne–Munford compactification.

2 The Stuck–Zimmer theorem

One of the most remarkable manifestations of rigidity for invariant random subgroups is the following celebrated result due to Stuck and Zimmer [1994].

Theorem 9. Let k be a local field and G be connected, simply connected semi-simple linear algebraic k-group. Assume that G has no k-anisotropic factors, has Kazhdan's property (T) and $\mathrm{rank}_k(G) \geq 2$. Then every properly ergodic and irreducible probability measure preserving Borel action of G is essentially free.

We recall that a probability measure preserving action is properly ergodic if it is ergodic and not essentially transitive, and is irreducible if every non-central normal subgroup is acting ergodically.

In the work of Stuck and Zimmer G is assumed to be a Lie group. The modifications necessary to deal with arbitrary local fields were carried on in Levit [2017b]. Much more generally, Bader and Shalom obtained a variant of the Stuck–Zimmer theorem for products of locally compact groups with property (T) in Bader and Shalom [2006].

The connection between invariant random subgroups and stabilizer structure for probability measure preserving actions allows one to derive the following:

Theorem 10. Let G be as in Theorem 9. Then any irreducible invariant random subgroup of G is either $\delta_{\{e\}}, \delta_G$ or μ_Γ for some irreducible lattice Γ in G.

We would like to point out that the Stuck–Zimmer theorem is a generalization of the following normal subgroup theorem of Margulis.

Theorem 11. Let G be as in Theorem 9 and Γ an irreducible lattice in G. Then any non-trivial normal subgroup $N \lhd \Gamma$ is either central or has finite index in Γ.

The Stuck–Zimmer theorem implies the normal subgroup theorem — indeed the ideas that go into its proof build upon the ideas of Margulis. One key ingredient is the intermediate factor theorem of Nevo and Zimmer, which in turn generalizes the factor theorem of Margulis. We point out that this aspect of the proof is entirely independent of property (T).

Question 2.1. Do Theorems 9 and 10 hold for all higher rank semisimple linear groups, regardless of property (T)?

Observe that the role played by Kazhdan's property (T) in the proof of Theorem 9 is in establishing the following fact.

Proposition 2.2. *Let G be a second countable locally compact group with Kazhdan's property (T). Then every properly ergodic probability measure preserving Borel action of G is not weakly amenable in the sense of Zimmer.*

The Nevo–Zimmer intermediate factor theorem is then used to show that a non-weakly amenable action is essentially free. On the other hand, the contrapositive of weak amenability follows quite readily from the fact that there are no non-trivial cocycles into amenable groups associated to probability measure preserving Borel actions of G.

To summarize, it is currently unknown if groups like $\mathrm{SL}_(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ or $\mathrm{SL}_2(\mathbb{Q}_p) \times \mathrm{SL}_2(\mathbb{Q}_p)$ admit any non-trivial irreducible invariant random subgroups not coming from lattices.

We would like to point out that rank one semisimple linear groups, discrete hyperbolic or relatively hyperbolic groups as well as mapping class groups and $\mathrm{Out}(F_n)$ have a large supply of exotic invariant random subgroups Bowen [2015b], Dahmani, Guirardel, and Osin [2017], and Bowen, Grigorchuk, and Kravchenko [2015].

3 The Benjamini–Schramm topology

Let $\mathfrak{M}$ be the space of all (isometry classes of) pointed proper metric spaces equipped with the Gromov–Hausdorff topology. This is a huge space and for many applications it is enough to consider compact subspaces of it obtained by bounding the geometry. That is, let $f(\epsilon, r)$ be an integer valued function defined on $(0,1) \times \mathbb{R}^{>0}$, and let $\mathfrak{M}_f$ consist of those spaces for which $\forall \epsilon, r$, the ϵ-entropy of the r-ball $B_X(r,p)$ around the special point is bounded by $f(\epsilon, r)$, i.e. no $f(\epsilon, r)+1$ points in $B_X(r,p)$ form an ϵ-discrete set. Then $\mathfrak{M}_f$ is a compact subspace of $\mathfrak{M}$.

In many situations one prefers to consider some variants of $\mathfrak{M}$ which carry more information about the spaces. For instance when considering graphs, it may be useful to add colors and orientations to the edges. The Gromov–Hausdorff distance defined on these objects should take into account the coloring and orientation. Another example is smooth Riemannian manifolds, in which case it is better to consider framed manifolds, i.e. manifold with a chosen point and a chosen frame at the tangent space at that point. In that case, one replace the Gromov–Hausdorff topology by the ones determined by (ϵ, r) relations (see Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a, Section 3] for details), which remembers also the directions from the special point.

We define the *Benjamini–Schramm* space $\mathcal{BS} = \mathrm{Prob}(\mathfrak{M})$ to be the space of all Borel probability measures on $\mathfrak{M}$ equipped with the weak-$*$ topology. Given f as above, we set $\mathcal{BS}_f := \mathrm{Prob}(\mathfrak{M}_f)$. Note that $\mathcal{BS}_f$ is compact.

The name of the space is chosen to hint that this is the same topology induced by 'local convergence', considered by Benjamini and Schramm in Benjamini and Schramm [2001], when restricting to measures on rooted graphs. Recall that a sequence of random rooted bounded degree graphs converges to a limiting distribution iff for every n the statistics of the n ball around the root (i.e. the probability vector corresponding to the finitely many possibilities for n-balls) converges to the limit.

The case of general proper metric spaces can be described similarly. A sequence $\mu_n \in \mathcal{BS}_f$ converges to a limit μ iff for any compact pointed 'test-space' $M \in \mathfrak{M}$, any r and some arbitrarily small[2] $\epsilon > 0$, the μ_n probability that the r ball around the special point is 'ϵ-close' to M tends to the μ-probability of the same event.

Example 3.1. An example of a point in $\mathcal{BS}$ is a measured metric space, i.e. a metric space with a Borel probability measure. A particular case is a finite volume Riemannian manifold — in which case we scale the Riemannian measure to be one, and then randomly choose a point and a frame.

Thus a finite volume locally symmetric space $M = \Gamma\backslash G/K$ produces both a point in the Benjamini–Schramm space and an IRS in G. This is a special case of a general analogy that I'll now describe. Given a symmetric space X, let us denote by $\mathfrak{M}(X)$ the space of all pointed (or framed) complete Riemannian orbifolds whose universal cover is X, and by $\mathcal{BS}(X) = \mathrm{Prob}(\mathfrak{M}(X))$ the corresponding subspace of the Benjamini–Schramm space.

Let G be a non-compact simple Lie group with maximal compact subgroup $K \leq G$ and an associated Riemannian symmetric space $X = G/K$. There is a natural map

$$\{\text{discrete subgroups of } G\} \to \mathfrak{M}(X),\ \Gamma \mapsto \Gamma\backslash X.$$

It can be shown that this map is continuous, hence inducing a continuous map

$$\mathrm{DIRS}(G) \to \mathcal{BS}(X).$$

[2] This doesn't mean that it happens for all ϵ.

It can be shown that the later map is one to one, and since $\mathrm{DIRS}(G)$ is compact, it is a homeomorphism to its image (see Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a, Corollary 3.4]). [3]

Remark 3.2 (Invariance under the geodesic flow)**.** Given a tangent vector $\overline{v}$ at the origin (the point corresponding to K) of $X = G/K$, define a map $\mathfrak{F}_{\overline{v}}$ from $\mathfrak{M}(X)$ to itself by moving the special point using the exponent of $\overline{v}$ and applying parallel transport to the frame. This induces a homeomorphism of $\mathfrak{BS}(X)$. The image of $\mathrm{DIRS}(G)$ under the map above is exactly the set of $\mu \in \mathfrak{BS}(X)$ which are invariant under $\mathfrak{F}_{\overline{v}}$ for all $\overline{v} \in T_K(G/K)$.

Thus we can view geodesic-flow invariant probability measures on framed locally X-manifolds as IRS on G and vice versa, and the Benjamini–Schramm topology on the first coincides with the IRS-topology on the second.

Remark 3.3. The analogy above can be generalised, to some extent, to the context of general locally compact groups. Given a locally compact group G, fixing a right invariant metric on G, we obtain a map $\mathrm{Sub}_G \to \mathfrak{M}$, $H \mapsto G/H$, where the metric on G/H is the induced one. Moreover, this map is continuous hence defines a continuous map $\mathrm{IRS}(G) \to \mathfrak{BS}$.

For the sake of simplicity let us forget 'the frame' and consider pointed X-manifolds, and $\mathfrak{BS}(X)$ as probability measures on such. We note that while for general Riemannian manifolds there is a benefit for working with framed manifolds, for locally symmetric spaces of non-compact type, pointed manifolds, and measures on such, behave nicely enough.

In order to examine convergence in $\mathfrak{BS}(X)$ it is enough to use as 'test-space' balls in locally X-manifolds. Moreover, since X is non-positively curved, a ball in an X-manifold is isometric to a ball in X iff it is contractible.

Note that since X is a homogeneous space, all choices of a probability measure on X correspond to the same point in $\mathfrak{BS}(X)$. Abusing notations, we shall denote this point by X.

Definition 3.4. Let us say that a sequence in $\mathfrak{BS}(X)$ is *Farber* if it converges to X.

For an X-manifold M and $r > 0$, we denote by $M_{\geq r}$ the r-thick part in M:

$$M_{\geq r} := \{x \in M : \mathrm{InjRad}_M(x) \geq r\},$$

where $\mathrm{InjRad}_M(x) = \sup\{\epsilon : B_M(x,\epsilon) \text{ is contractible}\}$.

[3] It was Miklos Abert who pointed out to me, about eight years ago, the analogy between Benjamini–Schramm convergence (at that time 'local convergence') and convergence of Invariant Random Subgroups, which later played an important rule in the work Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a] and earlier in the work of Abért, Y. Glasner, and Virág [2016, 2014].

Proposition 3.5. *Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a, Corollary 3.8] A sequence M_n of finite volume X-manifolds is Farber iff*

$$\frac{\mathrm{vol}((M_n)_{\geq r})}{\mathrm{vol}(M_n)} \to 1,$$

for every $r > 0$.

Theorem 1.12 can be reformulated as:

Theorem 3.6. *Let X be an irreducible Riemannian symmetric space of non compact type of rank at least 2. For any r and ϵ there is V such that if M is an X-manifold of volume $v \geq V$ then $\frac{\mathrm{vol}(M_{\geq r})}{v} \geq 1 - \epsilon$ (see Figure 2).*

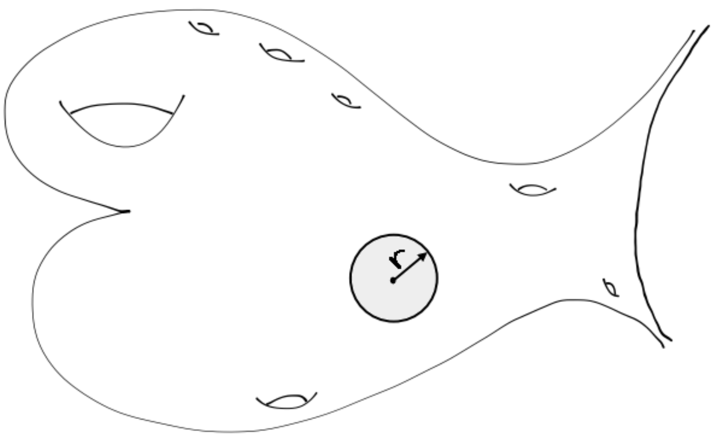


Figure 2: A large volume manifold is almost everywhere fat.

4 Applications to L_2-invariants

Let Γ be a uniform lattice in G. The right quasi-regular representation ρ_Γ of G in $L^2(\Gamma\backslash G, \mu_G)$ decomposes as a direct sum of irreducible representations. Every irreducible unitary representation π of G appears in ρ_Γ with finite multiplicity $m(\pi, \Gamma)$.

Definition 12. The *normalized relative Plancherel measure* of G with respect to Γ is an atomic measure on the unitary dual $\widehat{G}$ given by

$$\nu_\Gamma = \frac{1}{\mathrm{vol}(\Gamma\backslash G)} \sum_{\pi\in\widehat{G}} m(\pi,\Gamma)\delta_\pi .$$

The following result, extending earlier works of de George and Wallach [1978], Delorme [1986] and many others, was proved in Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a] for real Lie groups and then generalized to non-archimedean groups in Gelander and Levit [2017]:

Theorem 13. Let G be a semisimple analytic group in zero characteristic. Fix a Haar measure on G and let ν_G be the associated Plancherel measure on $\widehat{G}$. Let Γ_n be a uniformly discrete sequence of lattices in G with μ_{Γ_n} being weak-$*$ convergent to $\delta_{\{e\}}$. Then

$$\nu_{\Gamma_n}(E) \xrightarrow{n\to\infty} \nu_G(E)$$

for every relatively quasi-compact ν^G-regular subset $E \subset \widehat{G}$.

One of the consequence of Theorem 13 is the convergence of normalized Betti numbers (cf. Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2011]). Recently we were able to get rid of the co-compactness and the uniform discreteness assumptions and proved the following general version, making use of the Bowen–Elek Bowen [2015a, §4] simplicial approximation technique:

Theorem 4.1. *Abert, Bergeron, Biringer, and Gelander [n.d.] Let X be a symmetric space of non compact type of $dim(X) \neq 3$ and (M_n) is a weakly convergent sequence of finite volume X-manifolds. Then for all k, the normalized Betti numbers $b_k(M_n)/\mathrm{vol}(M_n)$ converge.*

Here, the only three-dimensional irreducible symmetric spaces of noncompact type are scales of $\mathbb{H}^3$. In fact, the conclusion of Theorem 4.1 is false when $X = \mathbb{H}^3$. As an example, let $K \subset S^3$ be a knot such that the complement $M = S^3 \setminus K$ admits a hyperbolic metric, e.g. the figure-8 knot. Using meridian–longitude coordinates, let M_n be obtained by Dehn filling M with slope $(1,n)$; then each M_n is a homology 3-sphere. The manifolds $M_n \to M$ geometrically, see Benedetti and Petronio [1992, Ch E.6], so the measures μ_{M_n} weakly converge to μ_M (c.f. Bader, Gelander, and Sauer [2016, Lemma 6.4]) and the volumes $\mathrm{vol}(M_n) \to \mathrm{vol}(M)$. However, $0 = b_1(M_n) \not\to b_1(M) = 1$, so the normalized Betti numbers of the sequence $M_1, M, M_2, M, \ldots$ do not converge.

Corollary 14. Suppose that (M_n) is a Farber sequence of finite volume X-manifolds. Then for all $k \in \mathbb{N}$, we have $b_k(M_n)/\mathrm{vol}(M_n) \to \beta_k^{(2)}(X)$.

In the thin case, we were able to push our analytic methods far enough to give a proof for $X = \mathbb{H}^d$, see Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2017a, Theorem 1.8]. Hence, there is no problem in allowing $X = \mathbb{H}^3$ in Corollary 14. The analog of Corollary 14 for p-adic Bruhat Tits buildings is proved in Gelander and Levit [2017].

5 Measures on the space of Riemannian manifolds

When $X = G/K$ is a symmetric space of noncompact type, say, the quotient of a discrete, torsion-free IRS Γ of G is a random X-manifold M. Fixing a base point p in X, the projection of p to $\Gamma\backslash X$ is a natural base point for the quotient. So, we can regard the quotient of an IRS as a random pointed X-manifold. In fact, the conjugation invariance of Γ directly corresponds to a property called *unimodularity* of the random pointed X-manifold, just as IRSs of discrete groups correspond to unimodular random Schreier graphs.

In Biringer and Abert [2016], the Abert and Biringer study unimodular probability measures on the more general space $\mathfrak{M}^d$ of all pointed Riemannian d-manifolds, equipped with the smooth topology. One can construct such unimodular measures from finite volume d-manifolds, or from IRSs of continuous groups as above (see Biringer and Abert [ibid., Proposition 1.9]). Under certain geometric assumptions like pinched negative curvature or local symmetry, they show that sequences of unimodular probability measures are precompact, in parallel with the compactness of the space of IRSs of a Lie group, see Biringer and Abert [ibid., Theorems 1.10 and 1.11]. They also show that unimodular measures on $\mathfrak{M}^d$ are just those that are 'compatible' with its foliated structure. Namely, $\mathfrak{M}^d$ is almost a foliated space, where a leaf is obtained by fixing a manifold M and varying the basepoint. While this foliation may be highly singular, they show in Biringer and Abert [ibid., Theorem 1.6] that after passing to an (actually) foliated desingularization, unimodular measures are just those that are created by integrating the Riemannian measures on the leaves against some invariant transverse measure. This is a precise analogue of the hard-to-formalize statement that a unimodular random graph is a random pointed graph in which the vertices are 'distributed uniformly' across each fixed graph.

6 Soficity of IRS

Definition 15. An IRS μ is *co-sofic* if it is a weak-$*$ limit in $\mathrm{IRS}(G)$ of ones supported on lattices.

The following result justify the name (cf. Abert, Gelander, and Nikolov [2017, Lemma 16]):

Proposition 6.1. *Let F_n be the free group of rank n. A Dirac mass δ_N, $N \lhd F_n$ is co-sofic iff the corresponding group $G = F_n/N$ is sofic.*

Given a group G it is natural to ask:

Question 6.2. Is every IRS in G co-sofic?

In particular for $G = F_n$ this is equivalent to the Aldous–Lyons conjecture that every unimodular network (supported on rank n Schreier graphs) is a limit of ones corresponding to finite Schreier graphs Aldous and Lyons [2007].

Therefore it is particularly intriguing to study Question 6.2 for G, a locally compact group admitting F_n as a lattice. This is the case for $G = SL_2(\mathbf{R}), SL_2(\mathbf{Q}_p)$ and $\mathrm{Aut}(T)$.

7 Exotic IRS

In the lack of Margulis' normal subgroup theorem there are IRS supported on non-lattices. Indeed, from a lattice $\Gamma \leq G$ and a normal subgroup of infinite index $N \lhd \Gamma$ one can cook an IRS in G supported on the closure of the conjugacy class N^G.

A more interesting example in $\mathrm{SO}(n,1)$ (from Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [2016]) is obtained by choosing two compact hyperbolic manifolds A, B with totally geodesic boundary, each with two components, and all four components are pairwise isometric and then glue random copies of A, B along an imaginary line to obtain a random hyperbolic manifold whose fundamental group is an IRS in $\mathrm{SO}(n,1)$. If A, B are chosen wisely, the random subgroup obtained is not contained in a lattice. However, all IRSs obtained that way are co-sofic. Other constructions of exotic IRS in $\mathrm{SO}(3,1)$ are given in Abert, Bergeron, Biringer, Gelander, Nikolov, Raimbault, and Samet [ibid.].

8 Existence

There are many well known examples of discrete groups without nontrivial IRS, for instance $\mathrm{PSL}_n(\mathbb{Q})$, and also the Tarski Monsters. In Gelander [2015, §8] I asked for non-discrete examples, and in particular weather the Neretin group (of almost adthomorphisms of a regular tree) admits non-trivial IRS. Recently Boudec and Bon [2017] constructed an example of a non-discrete locally compact group with no non-trivial IRS. (Note however that it is not compactly generated.)

9 Character Rigidity

Definition 16. Let Γ be a discrete group. A *character* on Γ is an irreducible positive definite complex-valued class function $\varphi : \Gamma \to \mathbb{C}$ satisfying $\varphi(e) = 1$.

The irreducibility of φ simply means that it cannot be written as a convex combination of two distinct characters. This notion was introduced by Thoma in Thoma [1964a,b]. In the abelian case Definition 16 reduces to the classical notion.

We will say that Γ has *Character rigidity* if only the obvious candidates occur as characters of Γ. The following theorem of Bekka [2007] is an outstanding example of such a result.

Theorem 17. Let φ be a character of the group $\Gamma = \mathrm{SL}_n(\mathbb{Z})$ for $n \geq 3$. Then either φ factors through an irreducible representation of some finite congruence quotient $\mathrm{SL}_n(\mathbb{Z}/N\mathbb{Z})$ or φ vanishes outside the center of Γ.

The connection between invariant random subgroups and characters arises from the following construction. Let (X, μ) be a Borel probability space with an action of Γ preserving μ. Consider the following real-valued function $\varphi : \Gamma \to \mathbf{R}$ that is associated to the action of Γ. The function g is given by

$$\varphi(\gamma) = \mu(\mathrm{Fix}(\gamma))$$

for every $\gamma \in \Gamma$, where

$$\mathrm{Fix}(\gamma) = \{x \in X \ : \ \gamma x = x\}.$$

For instance $\varphi(\gamma) = 1$ if γ lies in the kernel of the action and $\varphi(\gamma) = 0$ if μ-almost every point of X is not fixed by γ. It turns out that φ is a positive define class function satisfying $\varphi(e) = 1$.

Let Γ be an irreducible lattice in a higher rank semisimple linear group G with property (T). It can be shown by means of induced actions that Theorem 10 holds for the lattice Γ as well, namely any properly ergodic action of Γ has central stabilizers.

We see that Theorem 17 in fact implies Theorem 9 in the special case of the particular arithmetic group $\Gamma = \mathrm{SL}_n(\mathbb{Z}), n \geq 3$, which in turn implies the normal subgroup theorem of Margulis[4]. A character rigidity result is in general much stronger than invariant random subgroups rigidity — indeed, not all characters arise in the above manner from probability measure preserving actions.

Recently Peterson [2014] has been able to vastly generalize Bekka's result, as follows.

[4]The normal subgroup theorem for $\mathrm{SL}_n(\mathbb{Z})$ with $n \geq 3$ is in fact a much older theorem, dating back to Mennicke's work on the congruence subgroup problem Mennicke [1965].

Theorem 18. Character rigidity in the sense of Theorem 17 holds for any irreducible lattice in a higher rank semisimple Lie group without compact factors and with property (T).

Let us survey a few other well-known classification results for characters of discrete groups. In his original papers Thoma studied characters of the infinite symmetric group Thoma [1964a]. Dudko and Medynets studied characters of the Higman—Thompson and related groups Dudko and Medynets [2014]. Peterson and Thom establish character rigidity for linear groups over infinite fields or localizations of orders in number fields Peterson and Thom [2016b], generalizing several previous results Kirillov [1965] and Ovčinnikov [1971].

10 History

The interplay between a group theoretic and geometric viewpoints characterises the theory of IRS from its beginning. Two groundbreaking papers, Stuck and Zimmer [1994] and Aldous and Lyons [2007] represent these two points of view. Zimmer's work, throughout, was deeply influenced by Mackey's virtual group philosophy which draws an analogy between the subgroups of G and its ergodic actions. When G is a center free, higher rank simple Lie group, it is proved in Stuck and Zimmer [1994] that every non-essentially-free ergodic action is in fact a transitive action on the cosets of a lattice subgroup. These results can be viewed as yet another implementation of higher rank rigidity, but they also show that Mackey's analogy becomes much tighter when one considers non-essentially-free actions.

The Aldous–Lyons paper is influenced by the geometric notion of *Benjamini–Schramm* convergence in graphs, sometimes also referred to as *weak convergence* or as *convergence in local statistics*, developed in Aldous and Steele [2004],Benjamini and Schramm [2001],Benjamini, Lyons, Peres, and Schramm [1999]. Any finite graph[5] gives rise to a random rooted graph, upon choosing the root uniformly at random. Thus the collection of finite graphs, embeds as a discrete set, into the space of Borel probability measures on the (compact) space of rooted graphs. Random rooted graph in the w^*-closure of this set are subject to the *mass transport principal* introduced by Benjamini and Schramm [2001]: For every integrable function on the space of bi-rooted graphs

$$\int \sum_{x \in V(G)} f(G, o, x) d\mu([G, o]) = \int \sum_{x \in V(G)} f(G, x, o) d\mu([G, o]).$$

[5] or more generally an infinite graph whose automorphism group contains a lattice.

Aldous and Lyons define random unimodular graphs to be random rooted graphs subject to the mass transport principal. In Aldous and Lyons [2007, Question 10.1] they ask whether every random unimodular graph is in the w^*-closure of the set of finite graphs. When one specialises this theory to Schreier graphs of a given finitely generated group Γ (more generally to the quotients of the Cayley-Abels graph of a given compactly generated group G) one obtains the theory of IRS in Γ or in G. For probability measures on the Chabauty space of subgroups Sub_Γ — the mass transport principal is equivalent to invariance under the adjoint action of the group. When $\Gamma = F_d$ is the free group and $N \lhd \Gamma$ is a normal subgroup the group Γ/N is *sofic* in the sense of Gromov and Weiss if and only if the IRS δ_N is a w^*-limit of IRS supported on finite index subgroups. Thus the Aldous-Lyons question in the setting of Schreier graphs of F_d specializes to Gromov's question whether every group is sofic.

In a pair of papers Abért, Y. Glasner, and Virág [2016, 2014], Abért, Glasner and Virág introduced the notion of IRS and used it to answer a long standing question in graph theory. A sequence $\{X_n\}$ of finite, distinct d-regular Ramanujan graphs Benjamini–Schramm converges to the universal covering tree T_d. They provided a quantitative estimate for this result, for a Ramanujan graph X,

$$\Pr\{x \in X \mid \mathrm{inj}_X(x) \leq \beta \log\log(|X|)\} = O\left(\log(|X|)^{-\beta}\right),$$

where $\beta = (30\log(d-1))^{-1}$, $\mathrm{Inj}_X(x) = \max\{R \in \mathbf{N} \mid B_X(x,R) \text{ is contractible}\}$ and the probability is the uniform over the vertices of X. The proof combines the geometric and group theoretic viewpoints in an essential way: They start with a sequence of Ramanujan (Schreier) graphs $\{X_n\}$. Passing if necessary to a subsequence they assume that $X_n \to \Delta \backslash F_{d/2}$, where Δ is an IRS in $F_{d/2}$. Now the main technical result of their paper shows that the Schreier graph of an IRS has to satisfy Kesten's spectral gap theorem $\rho(\mathrm{Cay}(\Gamma/\Delta, S)) \geq \rho(\mathrm{Cay}(\Gamma, S))$ with equality if and only if $\Delta = \langle e \rangle$ a.s. Thus the limiting object is indeed the tree.

More generally they develop the theory of Benjamini–Shcramm limits of unimodular random graphs, as well as for Γ-Schreier graphs for arbitrary finitely generated group Γ. In this case the IRS version of Kesten's theorem reads $\rho(\mathrm{Cay}(\Gamma/\Delta, S)) \geq \rho(\mathrm{Cay}(\Gamma, S))$, with an (a.s.) equality, iff Δ is (a.s.) amenable. In hope of reproducing this same beautiful picture for general finitely generated groups, Abért, Glasner and Virág phrased a fundamental question that was quickly answered by Bader, Duchesne, and Lécureux [2016] giving rise to the following theorem: Every amenable IRS in a group Γ is supported on the subgroups of the amenable radical of Γ.

Independently of all of the above, Lewis Bowen in Bowen [2014], introduced the notion of an IRS, and of the *Poisson boundary relative to an IRS*. He used these notions to solve a long standing question in dynamics — proving that the Furstenberg entropy spectrum of the free group is a closed interval. Let (G, μ) be a locally compact group with a Borel probability measure on it, and (X, ν) a (G, μ) space. This means that $G \curvearrowright X$ acts on X measurably and ν is a μ-stationary probability measure in the sense that $\nu = \mu * \nu$. The Furstenberg entropy of this space is

$$h_\mu(X, \nu) = \int \int -\log \frac{d\eta \circ g}{d\eta}(x) d\eta(x) d\mu(g).$$

$\mathrm{Spec}(G, \mu) := \{h_\mu(X, \nu) \mid (X, \nu) \text{ an ergodic}(G, \mu)\text{-space}\}$ is called *the Furstenberg entropy spectrum* and it is bounded in the interval $[0, h_{max}(\mu)]$. The value 0 is obtained when the action is measure preserving, and the maximal value is always attained by the Poisson boundary $B(G, \mu)$. The study of the entropy spectrum is tightly related to the study of factors of the Poisson boundary. Nevo and Zimmer, Nevo and Zimmer [2000], consider a restricted spectrum, that comes only form actions subject to certain mixing properties and show that this restricted spectrum $\mathrm{Spec}'(G)$ is finite for a centre free, higher rank semisimple Lie group G. This result was then used in their proof of the intermediate factor theorem, which in retrospect also validated the proof of the Stuck-Zimmer theorem Nevo and Zimmer [1999, 2002b,a]. Bowen's work on IRS filled in a gap in the other direction — providing as it did many examples of stationary actions.

Let $K \in \mathrm{Sub}(G)$. The Poisson boundary $B(K \backslash G, \mu)$ is the (Borel) quotient of the space of all μ-random walks on $K \backslash G$ under the shift $\sigma(Kg_0, Kg_1, Kg_2, \ldots) = (Kg_1, Kg_2, \ldots)$. If $\Delta = N$ happens to be normal then one retains the Kaimanovich–Vershik description of the Poisson boundary on G/N and clearly G acts on this space from the left giving it the structure of a (G, μ)-space. In the more general setting introduced by Bowen, G still acts on the natural bundle over $\mathrm{Sub}(G)$, where the fibre over $K \in \mathrm{Sub}(G)$ is $B(K \backslash G, \mu)$. The natural action of G on the space of all walks on all these coset spaces, given by $g(Kg_1, Kg_2, \ldots) = (gKg^{-1}gg_1, gKg^{-1}gg_2, \ldots)$ clearly commutes with the shift and gives rise to a well defined action of (G, μ) on this bundle. Any choice of an IRS $\theta \in \mathrm{IRS}(G)$ gives rise to a (G, μ)-stationary measure on this bundle. Now for $F_d = \langle s_1, \ldots, s_d \rangle$, $\mu = \frac{1}{2d}\left(\sum_{i=1}^d s_i + s_i^{-1}\right)$ the proof that $\mathrm{Spec}(F_d, \mu) = [0, h_{\max}(\mu)]$ is completed by finding a certain path $\alpha : I :\to \mathrm{IRS}(F_d)$ in the space of IRSs with the following properties: (i) the path starts at the trivial IRS (corresponding to the action on the Poisson boundary), (ii) it ends at an IRS giving rise to arbitrarily small entropy values and (iii) the entropy function is continuous on this path. Continuity of the entropy function is very special and so are the IRS that are chosen in order to allow for this continuity. The existence of paths on the other hand is actually general, in Bowen [2015b] Bowen proves that the collection of ergodic IRS on F_d that are not supported on finite index subgroups is path connected.

Vershik [2012, 2010], also indpendently, arrived at IRS from his study of the representation theory and especially the characters of S_f^∞ — the group of finitely supported permutations of a countable set. To an IRS $\mu \in \mathrm{IRS}(\Gamma)$ in a countable group define its *Vershik character* as follows

$$\phi_\mu : \Gamma \to \mathbf{R}_{\geq 0}, \qquad \phi_\mu(\gamma) = \mu\left(\{\Delta \in \mathrm{Sub}(\Gamma) \mid \gamma \in \Delta\}\right).$$

If the IRS is realized as the stabilizer Γ_x of a random point in a p.m.p. action $\Gamma \curvearrowright (X, \nu)$ (by Abért, Y. Glasner, and Virág [2014], every IRS can be realized in this fashion), the same IRS is given by $\phi_\nu(\gamma) = \nu(Fix(\gamma))$. Vershik also describes the GNS constuctions associated with this character. Let $R = \{(x, y) \in X \times X | y \in \Gamma x\}$ and let η be the infinite measure on R given by $\int f(x, y)\eta(x, y) = \int \sum_{y \in \Gamma x} f(x, y) d\mu(x)$. Γ acts on R via its action on the first coordinate $\gamma(x, y) = (\gamma x, y)$ and hence it acts on the Hilbert space $L_2(R, \eta)$. Let $\chi(x, y) = 1_{x=y} \in L_2(R, \eta)$ be the characteristic function of the diagonal. It is easy to verify that $phi_\mu(\gamma) = \langle \gamma\chi, \chi \rangle$. The defintion of the Vershik character clarified the deep connection between character rigidity in the snese of Connes and and the Stuck–Zimmer theorem.

Acknowledgments. Several people helped me in writing up parts of this paper: Ian Biringer with Section 5, Arie Levit with Section 9 and Section 2 and Yair Glasner with Section 10.

References

M. Abert, N. Bergeron, I. Biringer, and T. Gelander (n.d.). "Convergence of normalized Betti numbers in nonpositive curvature" (cit. on p. 1353).

Miklos Abert, Nicolas Bergeron, Ian Biringer, Tsachik Gelander, Nikolay Nikolov, Jean Raimbault, and Iddo Samet (2011). "On the growth of Betti numbers of locally symmetric spaces". *C. R. Math. Acad. Sci. Paris* 349.15-16, pp. 831–835. MR: 2835886 (cit. on p. 1353).

– (Dec. 2016). "On the growth of L2-invariants of locally symmetric spaces, II: exotic invariant random subgroups in rank one". arXiv: 1612.09510 (cit. on p. 1355).

– (2017a). "On the growth of L^2-invariants for sequences of lattices in Lie groups". *Ann. of Math. (2)* 185.3, pp. 711–790. arXiv: 1210.2961. MR: 3664810 (cit. on pp. 1339, 1340, 1344, 1347, 1350–1354).

– (2017b). "On the growth of L^2-invariants for sequences of lattices in Lie groups". *Ann. of Math. (2)* 185.3, pp. 711–790. MR: 3664810.

Miklos Abert, Tsachik Gelander, and Nikolay Nikolov (2017). "Rank, combinatorial cost, and homology torsion growth in higher rank lattices". *Duke Math. J.* 166.15, pp. 2925–2964. MR: 3712168 (cit. on p. 1354).

Miklós Abért, Yair Glasner, and Bálint Virág (2014). "Kesten's theorem for invariant random subgroups". *Duke Math. J.* 163.3, pp. 465–488. MR: 3165420 (cit. on pp. 1351, 1358, 1360).

– (2016). "The measurable Kesten theorem". *Ann. Probab.* 44.3, pp. 1601–1646. MR: 3502591 (cit. on pp. 1351, 1358).

David Aldous and Russell Lyons (2007). "Processes on unimodular random networks". *Electron. J. Probab.* 12, no. 54, 1454–1508. MR: 2354165 (cit. on pp. 1355, 1357, 1358).

David Aldous and J. Michael Steele (2004). "The objective method: probabilistic combinatorial optimization and local weak convergence". In: *Probability on discrete structures*. Vol. 110. Encyclopaedia Math. Sci. Springer, Berlin, pp. 1–72. MR: 2023650 (cit. on p. 1357).

Uri Bader, Pierre-Emmanuel Caprace, Tsachik Gelander, and Shahar Mozes (2012). "Simple groups without lattices". *Bull. Lond. Math. Soc.* 44.1, pp. 55–67. MR: 2881324.

Uri Bader, Bruno Duchesne, and Jean Lécureux (2016). "Amenable invariant random subgroups". *Israel J. Math.* 213.1. With an appendix by Phillip Wesolek, pp. 399–422. MR: 3509477 (cit. on p. 1358).

Uri Bader, Tsachik Gelander, and Roman Sauer (Dec. 2016). "Homology and homotopy complexity in negative curvature". arXiv: 1612.04871 (cit. on p. 1353).

Uri Bader and Yehuda Shalom (2006). "Factor and normal subgroup theorems for lattices in products of groups". *Invent. Math.* 163.2, pp. 415–454. MR: 2207022 (cit. on p. 1348).

Y. Barnea and M. Larsen (2004). "Random generation in semisimple algebraic groups over local fields". *J. Algebra* 271.1, pp. 1–10. MR: 2022476 (cit. on p. 1345).

Bachir Bekka (2007). "Operator-algebraic superridigity for $\mathrm{SL}_n(\mathbb{Z}), n \geq 3$". *Invent. Math.* 169.2, pp. 401–425. MR: 2318561 (cit. on p. 1356).

Riccardo Benedetti and Carlo Petronio (1992). *Lectures on hyperbolic geometry*. Universitext. Springer-Verlag, Berlin, pp. xiv+330. MR: 1219310 (cit. on p. 1353).

I. Benjamini, R. Lyons, Y. Peres, and O. Schramm (1999). "Group-invariant percolation on graphs". *Geom. Funct. Anal.* 9.1, pp. 29–66. MR: 1675890 (cit. on p. 1357).

Itai Benjamini and Oded Schramm (2001). "Recurrence of distributional limits of finite planar graphs". *Electron. J. Probab.* 6, no. 23, 13. MR: 1873300 (cit. on pp. 1350, 1357).

Ian Biringer and Miklos Abert (June 2016). "Unimodular measures on the space of all Riemannian manifolds". arXiv: 1606.03360 (cit. on p. 1354).

Adrien Le Boudec and Nicolas Matte Bon (Sept. 2017). "Locally compact groups whose ergodic or minimal actions are all free". arXiv: 1709.06733 (cit. on p. 1355).

Lewis Bowen (2014). "Random walks on random coset spaces with applications to Furstenberg entropy". *Invent. Math.* 196.2, pp. 485–510. MR: 3193754 (cit. on p. 1358).

– (2015a). "Cheeger constants and L^2-Betti numbers". *Duke Math. J.* 164.3, pp. 569–615. MR: 3314481 (cit. on p. 1353).

– (2015b). "Invariant random subgroups of the free group". *Groups Geom. Dyn.* 9.3, pp. 891–916. MR: 3420547 (cit. on pp. 1349, 1359).

Lewis Bowen, Rostislav Grigorchuk, and Rostyslav Kravchenko (2015). "Invariant random subgroups of lamplighter groups". *Israel J. Math.* 207.2, pp. 763–782. MR: 3359717 (cit. on p. 1349).

F. Dahmani, V. Guirardel, and D. Osin (2017). "Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces". *Mem. Amer. Math. Soc.* 245.1156, pp. v+152. arXiv: 1111.7048. MR: 3589159 (cit. on p. 1349).

Patrick Delorme (1986). "Formules limites et formules asymptotiques pour les multiplicités dans $L^2(G/\Gamma)$". *Duke Math. J.* 53.3, pp. 691–731. MR: 860667 (cit. on p. 1353).

Artem Dudko and Konstantin Medynets (2014). "Finite factor representations of Higman-Thompson groups". *Groups Geom. Dyn.* 8.2, pp. 375–389. arXiv: 1212.1230. MR: 3231220 (cit. on p. 1357).

Mikolaj Fraczyk (Dec. 2016). "Strong Limit Multiplicity for arithmetic hyperbolic surfaces and 3-manifolds". arXiv: 1612.05354 (cit. on p. 1347).

Harry Furstenberg (1976). "A note on Borel's density theorem". *Proc. Amer. Math. Soc.* 55.1, pp. 209–212. MR: 0422497 (cit. on p. 1340).

Tsachik Gelander (2014). "Lectures on lattices and locally symmetric spaces". In: *Geometric group theory*. Vol. 21. IAS/Park City Math. Ser. Amer. Math. Soc., Providence, RI, pp. 249–282. MR: 3329730.

– (Mar. 2015). "A lecture on Invariant Random Subgroups". arXiv: 1503.08402 (cit. on p. 1355).

– (2018). "Kazhdan–Margulis theorem for invariant random subgroups". *Adv. Math.* 327, pp. 47–51. MR: 3761990 (cit. on p. 1345).

Tsachik Gelander and Arie Levit (May 2016). "Local Rigidity Of Uniform Lattices". arXiv: 1605.01693 (cit. on p. 1347).

– (July 2017). "Invariant random subgroups over non-Archimedean local fields". arXiv: 1707.03578 (cit. on pp. 1341, 1345, 1347, 1353, 1354).

David L. de George and Nolan R. Wallach (1978). "Limit formulas for multiplicities in $L^2(\Gamma\backslash G)$". *Ann. of Math. (2)* 107.1, pp. 133–150. MR: 0492077 (cit. on p. 1353).

E. Glasner and B. Weiss (1997). "Kazhdan's property T and the geometry of the collection of invariant measures". *Geom. Funct. Anal.* 7.5, pp. 917–935. MR: 1475550 (cit. on p. 1344).

Yair Glasner (2017). "Invariant random subgroups of linear groups". *Israel J. Math.* 219.1. With an appendix by Tsachik Gelander and Glasner, pp. 215–270. arXiv: 1407.2872. MR: 3642021.

D. A. Každan (1967). "On the connection of the dual space of a group with the structure of its closed subgroups". *Funkcional. Anal. i Priložen.* 1, pp. 71–74. MR: 0209390 (cit. on p. 1344).

D. A. Každan and G. A. Margulis (1968). "A proof of Selberg's hypothesis". *Mat. Sb. (N.S.)* 75 (117), pp. 163–168. MR: 0223487 (cit. on p. 1342).

A. A. Kirillov (1965). "Positive-definite functions on a group of matrices with elements from a discrete field". *Dokl. Akad. Nauk SSSR* 162, pp. 503–505. MR: 0193183 (cit. on p. 1357).

Benoît Kloeckner (2009). "The space of closed subgroups of $\mathbb{R}^n$ is stratified and simply connected". *J. Topol.* 2.3, pp. 570–588. MR: 2546586.

Masatake Kuranishi (1951). "On everywhere dense imbedding of free groups in Lie groups". *Nagoya Math. J.* 2, pp. 63–71. MR: 0041145 (cit. on p. 1340).

A. Levit (n.d.). "Weak uniform discreteness of lattices in positive characteristic" (cit. on p. 1346).

Arie Levit (May 2017a). "On Benjamini–Schramm limits of congruence subgroups". arXiv: 1705.04200 (cit. on p. 1347).

– (2017b). "The Nevo-Zimmer intermediate factor theorem over local fields". *Geom. Dedicata* 186, pp. 149–171. MR: 3602889 (cit. on p. 1348).

G. A. Margulis (1991). *Discrete subgroups of semisimple Lie groups*. Vol. 17. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin, pp. x+388. MR: 1090825 (cit. on pp. 1342, 1344).

Jens L. Mennicke (1965). "Finite factor groups of the unimodular group". *Ann. of Math. (2)* 81, pp. 31–37. MR: 0171856 (cit. on p. 1356).

Amos Nevo and Robert J. Zimmer (1999). "Homogenous projective factors for actions of semi-simple Lie groups". *Invent. Math.* 138.2, pp. 229–252. MR: 1720183 (cit. on p. 1359).

– (2000). "Rigidity of Furstenberg entropy for semisimple Lie group actions". *Ann. Sci. École Norm. Sup. (4)* 33.3, pp. 321–343. MR: 1775184 (cit. on p. 1359).

– (2002a). "A generalization of the intermediate factors theorem". *J. Anal. Math.* 86, pp. 93–104. MR: 1894478 (cit. on p. 1359).

– (2002b). "A structure theorem for actions of semisimple Lie groups". *Ann. of Math. (2)* 156.2, pp. 565–594. MR: 1933077 (cit. on p. 1359).

S. V. Ovčinnikov (1971). “Positive definite functions on Chevalley groups”. *Funkcional. Anal. i Priložen.* 5.1, pp. 91–92. MR: 0291371 (cit. on p. 1357).

Jesse Peterson (2014). “Character rigidity for lattices in higher-rank groups” (cit. on p. 1356).

Jesse Peterson and Andreas Thom (2016a). “Character rigidity for special linear groups”. *J. Reine Angew. Math.* 716, pp. 207–228. MR: 3518376.

– (2016b). “Character rigidity for special linear groups”. *J. Reine Angew. Math.* 716, pp. 207–228. MR: 3518376 (cit. on p. 1357).

Vladimir Platonov and Andrei Rapinchuk (1994). *Algebraic groups and number theory*. Vol. 139. Pure and Applied Mathematics. Translated from the 1991 Russian original by Rachel Rowen. Academic Press, Inc., Boston, MA, pp. xii+614. MR: 1278263.

M. S. Raghunathan (1972). *Discrete subgroups of Lie groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68. Springer-Verlag, New York-Heidelberg, pp. ix+227. MR: 0507234 (cit. on p. 1346).

Jean Raimbault (Nov. 2013). “On the convergence of arithmetic orbifolds”. arXiv: 1311.5375 (cit. on p. 1347).

Alireza Salehi Golsefidy (2013). “Lattices of minimum covolume are non-uniform”. *Israel J. Math.* 196.1, pp. 363–373. MR: 3096596 (cit. on p. 1346).

Jean-Pierre Serre (2006). *Lie algebras and Lie groups*. Vol. 1500. Lecture Notes in Mathematics. 1964 lectures given at Harvard University, Corrected fifth printing of the second (1992) edition. Springer-Verlag, Berlin, pp. viii+168. MR: 2179691 (cit. on p. 1344).

Garrett Stuck and Robert J. Zimmer (1994). “Stabilizers for ergodic actions of higher rank semisimple groups”. *Ann. of Math. (2)* 139.3, pp. 723–747. MR: 1283875 (cit. on pp. 1348, 1357).

Elmar Thoma (1964a). “Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe”. *Math. Z.* 85, pp. 40–61. MR: 0173169 (cit. on pp. 1356, 1357).

– (1964b). “Über unitäre Darstellungen abzählbarer, diskreter Gruppen”. *Math. Ann.* 153, pp. 111–138. MR: 0160118 (cit. on p. 1356).

Hiraku Toyama (1949). “On discrete subgroups of a Lie group”. *Kōdai Math. Sem. Rep.,* 1.2. Volume numbers not printed on issues until Vol. 7 (1955), pp. 36–37. MR: 0029918 (cit. on p. 1340).

A. M. Vershik (2010). “Nonfree actions of countable groups and their characters”. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 378.Teoriya Predstavleniĭ, Dinamicheskie Sistemy, Kombinatornye Metody. XVIII, pp. 5–16, 228. MR: 2749291 (cit. on p. 1360).

– (2012). “Totally nonfree actions and the infinite symmetric group”. *Mosc. Math. J.* 12.1, pp. 193–212, 216. MR: 2952431 (cit. on p. 1360).

Robert J. Zimmer (1984). *Ergodic theory and semisimple groups*. Vol. 81. Monographs in Mathematics. Birkhäuser Verlag, Basel, pp. x+209. MR: 776417.

Received 2017-11-27.

Tsachik Gelander
tsachik.gelander@weizmann.ac.il

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1363–1384)

SOME RESULTS ON AFFINE DELIGNE–LUSZTIG VARIETIES

Xuhua He (何旭华)

Abstract

The study of affine Deligne-Lusztig varieties originally arose from arithmetic geometry, but many problems on affine Deligne-Lusztig varieties are purely Lie-theoretic in nature. This survey deals with recent progress on several important problems on affine Deligne-Lusztig varieties. The emphasis is on the Lie-theoretic aspect, while some connections and applications to arithmetic geometry will also be mentioned.

1 Introduction

1.1 Bruhat decomposition and conjugacy classes. Let $\mathbb{G}$ be a connected reductive group over a field $\mathbf{k}$ and $G = \mathbb{G}(\mathbf{k})$. In this subsection, we assume that $\mathbf{k}$ is algebraically closed. Let B be a Borel subgroup of G and W be the finite Weyl group of G. The Bruhat decomposition $G = \sqcup_{w \in W} BwB$ plays a fundamental role in Lie theory. This is explained by Lusztig [2010] in the memorial conference of Bruhat:

"By allowing one to reduce many questions about G to questions about the Weyl group W, Bruhat decomposition is indispensable for the understanding of both the structure and representations of G."

Below we mention two examples of the interaction between the Bruhat decomposition and the (ordinary and twisted) conjugation action of G.

1. Assume that $\mathbf{k} = \overline{\mathbb{F}}_q$ and σ is the Frobenius of $\mathbf{k}$ over $\mathbb{F}_q$. We assume that $\mathbb{G}$ is defined over $\mathbb{F}_q$ and we denote by σ the corresponding Frobenius morphism on G. The (classical) Deligne-Lusztig varieties was introduced by Deligne and Lusztig in their seminal work Deligne and Lusztig [1976]. For any element $w \in W$, the corresponding Deligne-Lusztig variety X_w is a subvariety of the flag variety G/B defined by

$$X_w = \{gB \in G/B; g^{-1}\sigma(g) \in BwB\}.$$

X. H. was partially supported by NSF DMS-1463852.
MSC2010: primary 14L05; secondary 20G25.
Keywords: Affine Deligne-Lusztig varieties, loop groups, affine Weyl groups.

By Lang's theorem, the variety X_w is always nonempty. It is a locally closed, smooth variety of dimension $\ell(w)$. The finite reductive group $\mathbb{G}(\mathbb{F}_q)$ acts naturally on X_w and on the cohomology of X_w. The Deligne-Lusztig variety X_w plays a crucial role in the representation theory of finite reductive groups, see Deligne and Lusztig [1976] and Lusztig [1984]. The structure of X_w has also found important applications in number theory, e.g., in the work of Rapoport, Terstiege, and W. Zhang [2013], and in the work of Li and Y. Zhu [2017] on the proof of special cases of the "arithmetic fundamental lemma" of W. Zhang [2012].

2. Let $\mathbf{k}$ be any algebraically closed field. In a series of papers Lusztig [2012], Lusztig discovered a deep relation between the unipotent conjugacy classes of G and the conjugacy classes of W, via the study of the intersection of the unipotent conjugacy classes with the Bruhat cells of G.

1.2 Affine Deligne-Lusztig varieties. The main objects of this survey are affine Deligne-Lusztig varieties, analogous of classical Deligne-Lusztig varieties for loop groups.

Unless otherwise stated, in the rest of this survey we assume that $\mathbf{k} = \overline{\mathbb{F}}_q((\epsilon))$. Let σ be the Frobenius morphism of $\mathbf{k}$ over $\mathbb{F}_q((\epsilon))$. We assume that $\mathbb{G}$ is defined over $\mathbb{F}_q((\epsilon))$ and we denote by σ the corresponding Frobenius morphism on the loop group $G = \mathbb{G}(\mathbf{k})$. We choose a σ-stable Iwahori subgroup I of G. If G is unramified, then we also choose a σ-stable hyperspecial parahoric subgroup $K \supset I$. The affine flag variety $Fl = G/I$ and the affine Grassmannian $Gr = G/K$ (if G is unramified) have natural scheme structures.[1]

Let S be a maximal $\mathbf{k}$-split torus of G defined over $\mathbb{F}_q((\epsilon))$ and let T be its centralizer, a maximal torus of G. The Iwahori-Weyl group associated to S is

$$\tilde{W} = N(\mathbf{k})/T(\mathbf{k})_1,$$

where N is the normalizer of S in G and $T(\mathbf{k})_1$ is the maximal open compact subgroup of $T(\mathbf{k})$. The group $\tilde{W}$ is also a split extension of the relative (finite) Weyl group W_0 by the normal subgroup $X_*(T)_{\Gamma_0}$, where $X_*(T)$ is the coweight lattice of T and Γ_0 is the Galois group of $\overline{\mathbf{k}}$ over $\mathbf{k}$ (cf. Pappas and Rapoport [2008, Appendix]). The group $\tilde{W}$ has a natural quasi-Coxeter structure. We denote by ℓ and $\leqslant$ the length function and the Bruhat order on $\tilde{W}$. We have the following generalization of the Bruhat decomposition

$$G = \sqcup_{w \in \tilde{W}} IwI,$$

[1] One may replace $\overline{\mathbb{F}}_q((\epsilon))$ by the fraction field of the Witt ring. In that case, the affine Grassmannian Gr and the affine flag variety Fl have the structure of perfect schemes, thanks to the recent breakthrough of X. Zhu [2017], and of Bhatt and Scholze [2017]. Many of the results we discuss in this survey hold for the fraction field of the Witt ring as well.

due to Iwahori and Matsumoto [1965] in the split case, and to Bruhat and Tits [1972] in the general case. If G is unramified, then we also have

$$G = \sqcup_{\lambda \text{ is a dominant coweight}} K\epsilon^{\lambda} K.$$

Affine Deligne-Lusztig varieties were introduced by Rapoport in Rapoport [2005]. Compared to the classical Deligne-Lusztig varieties, we need two parameters here: an element w in the Iwahori-Weyl group W and an element b in the loop group G. The corresponding affine Deligne-Lusztig variety (in the affine flag variety) is defined as

$$X_w(b) = \{gI \in G/I; g^{-1} b\sigma(g) \in IwI\} \subset Fl.$$

If G is unramified, one may use a dominant coweight λ instead of an element in $\tilde{W}$ and define the affine Deligne-Lusztig variety (in the affine Grassmannian) by

$$X_\lambda(b) = \{gK \in G/K; g^{-1} b\sigma(g) \in K\epsilon^{\lambda} K\} \subset Gr.$$

Affine Deligne-Lusztig varieties are schemes locally of finite type over $\overline{\mathbb{F}}_q$. Also the varieties are isomorphic if the element b is replaced by another element b' in the same σ-conjugacy class.

A major difference between affine Deligne-Lusztig varieties and classical Deligne-Lusztig varieties is that affine Deligne-Lusztig varieties have the second parameter: the element b, or the σ-conjugacy class $[b]$ in the loop group G; while in the classical case considered in Section 1.1, by Lang's theorem there is only one σ-conjugacy class in $\mathbb{G}(\overline{\mathbb{F}}_q)$ and thus adding a parameter $b \in \mathbb{G}(\overline{\mathbb{F}}_q)$ does not give any new variety.

The second parameter $[b]$ in the affine Deligne-Lusztig varieties makes them rather challenging to study, both from the Lie-theoretic point of view, and from the arithmetic-geometric point of view. Below we list some major problems on the affine Deligne-Lusztig varieties:

- When is an affine Deligne-Lusztig variety nonempty?
- If it is nonempty, what is its dimension?
- What are the connected components?
- Is there a simple geometric structure for certain affine Deligne-Lusztig varieties?

We may also consider the affine Deligne-Lusztig varieties associated to arbitrary parahoric subgroups, besides hyperspecial subgroups and Iwahori subgroups. This will be discussed in Section 7.

1.3 A short overview of $X(\mu, b)$. The above questions may also be asked for a certain union $X(\mu, b)$ of affine Deligne-Lusztig varieties in the affine flag variety.

Let μ be a dominant coweight of G with respect to a given Borel subgroup of G over **k** (in applications to number theory, μ usually comes from a Shimura datum). The admissible set $\mathrm{Adm}(\mu)$ was introduced by Kottwitz and Rapoport in R. Kottwitz and Rapoport [2000]. It is defined by

$$\mathrm{Adm}(\mu) = \{w \in \tilde{W}; w \leqslant t^{x(\mu)} \text{ for some } x \in W_0\}.$$

We may explain it in a more Lie-theoretic language. Let $Gr_{\mathcal{G}}$ be the deformation from the affine Grassmannian to the affine flag variety Gaitsgory [2001]. The coherence conjecture of Pappas and Rapoport [2008] implies that the special fiber of the global Schubert variety $\overline{Gr}_{\mathcal{G},\mu}$ associated to the coweight μ (cf. X. Zhu [2014, Definition 3.1]) is $\cup_{w \in \mathrm{Adm}(\mu)} IwI/I$. This conjecture was proved by Zhu in X. Zhu [ibid.]. Now we set

$$X(\mu, b) = \cup_{w \in \mathrm{Adm}(\mu)} X_w(b) \subset Fl.$$

This is a closed subscheme of Fl and serves as the group-theoretic model for the Newton stratum corresponding to $[b]$ in the special fiber of a Shimura variety giving rise to the datum (G, μ).

It is also worth mentioning that, although the admissible set $\mathrm{Adm}(\mu)$ has a rather simple definition, it is a very complicated combinatorial object. We refer to the work of Haines and Ngô Haines and Châu [2002], and the recent joint work of the author with Haines and He [2017] for some properties of $\mathrm{Adm}(\mu)$.

1.4 Current status. Affine Deligne-Lusztig varieties in the affine Grassmannian are relatively more accessible than the ones in the affine flag variety, mainly due to the following two reasons:

- The set of dominant coweights is easier to understand than the Iwahori-Weyl group;
- For $X_\lambda(b)$ the group G is unramified while for $X_w(b)$, we need to deal with ramified, or even non quasi-split reductive groups.

For an unramified group G, we also have the fibration $\cup_{w \in W_0 t^\lambda W_0} X_w(b) \to X_\lambda(b)$, with fibers isomorphic to the flag variety of $\mathbb{G}(\overline{\mathbb{F}}_q)$. Thus much information on $X_\lambda(b)$ can be deduced from $X_w(b)$.

Nevertheless, the study of the affine Deligne-Lusztig varieties in affine Grassmannian is a very challenging task and has attracted the attention of experts in arithmetic geometry in the past two decades. It is a major achievement in arithmetic geometry to obtain a fairly good understanding on these varieties.

As to the affine Deligne-Lusztig varieties in the affine flag varieties, the situation is even more intriguing. We have made significant progress in the past 10 years in this direction, yet many aspects of $X_w(b)$ remain rather mysterious. I hope that by combining various Lie-theoretic methods together with arithmetic-geometric methods, our knowledge on affine Deligne-Lusztig varieties will be considerably advanced.

In the rest of the survey, we will report on some recent progress on the affine Deligne-Lusztig varieties.

Acknowledgments. We thank Ulrich Görtz, Urs Hartl, George Lusztig, Michael Rapoport, Sian Nie and Rong Zhou for useful comments.

2 Some relation with affine Hecke algebras

2.1 The set $B(G)$ and Kottwitz's classification. Let $B(G)$ be the set of σ-conjugacy classes of G. R. E. Kottwitz [1985] and R. E. Kottwitz [1997] gave a classification of the set $B(G)$, generalizing the Dieudonné-Manin classification of isocrystals by their Newton polygons. Any σ-conjugacy class $[b]$ is determined by two invariants:

- The element $\kappa([b]) \in \pi_1(G)_\Gamma$, where Γ is the Galois group of $\bar{\mathbf{k}}$ over $\mathbb{F}_q((\epsilon))$;
- The Newton point ν_b in the dominant chamber of $X_*(T)_{\Gamma_0} \otimes \mathbb{Q}$.

A different point of view, which is quite useful in this survey, is the relation between the set $B(G)$ with the set $B(\tilde{W}, \sigma)$ of σ-conjugacy classes of $\tilde{W}$. Recall that $\tilde{W} = N(\mathbf{k})/T(\mathbf{k})_1$. The natural embedding $N(\mathbf{k}) \to G$ induces a natural map $\Psi : B(\tilde{W}, \sigma) \to B(G)$. By Görtz, Haines, R. E. Kottwitz, and Reuman [2010] and He [2014], the map Ψ is surjective. The map Ψ is not injective. However, there exists an important family $B(\tilde{W}, \sigma)_{str}$ of straight σ-conjugacy classes of $\tilde{W}$. By definition, a σ-conjugacy class $\mathcal{O}$ of $\tilde{W}$ is *straight* if it contains an element $w \in \mathcal{O}$ such that $\ell(w\sigma(w) \cdots \sigma^{n-1}(w)) = n\ell(w)$ for all $n \in \mathbb{N}$. The following result is discovered in He [ibid., Theorem 3.7].

Theorem 2.1. *The map $\Psi : B(\tilde{W}, \sigma) \to B(G)$ induces a bijection*

$$B(\tilde{W}, \sigma)_{str} \longleftrightarrow B(G).$$

This result gives the parametrization of the σ-conjugacy classes of G in terms of the set of straight σ-conjugacy classes of its Iwahori-Weyl group $\tilde{W}$. In particular, the two parameters occurring in the definition of the affine Deligne-Lusztig variety $X_w(b)$ are all from $\tilde{W}$.

Note that the affine Deligne-Lusztig variety $X_w(b)$ is closely related to the intersection $IwI \cap [b]$. This intersection is very complicated in general. However, it is discovered in He [2014] that for certain elements $w \in \tilde{W}$, the intersection $IwI \cap [b]$ equals IwI. More precisely, we denote by $\tilde{W}_{\sigma-\min}$ the set of elements in $\tilde{W}$ that are of minimal length in their σ-conjugacy classes. Then

$$\text{For } w \in \tilde{W}_{\sigma-\min},\ IwI \subset [b] \text{ if } [b] = \Psi(w).$$

This serves as the starting point of the reduction method for affine Deligne-Lusztig varieties $X_w(b)$ for arbitrary w.

2.2 "Dimension=Degree" theorem. Deligne and Lusztig introduced in Deligne and Lusztig [1976] a reduction method to study the classical Deligne-Lusztig varieties. Their method works for the affine Deligne-Lusztig varieties as well. Some combinatorial properties of affine Weyl groups established in joint work with Nie 2014 allow us to reduce the study of $X_w(b)$ for any w, via the reduction method à la Deligne and Lusztig, to the study of $X_w(b)$ for $w \in \tilde{W}_{\sigma-\min}$.

The explicit reduction procedure, however, is rather difficult to keep track of. In He [2014], we discovered that the same reduction procedure appears in a totally different context as follows.

Let H be the affine Hecke algebra (over $\mathbb{Z}[v^{\pm 1}]$) associated to $\tilde{W}$. Let $[\tilde{H}, \tilde{H}]_\sigma$ be the σ-twisted commutator, i.e. the $\mathbb{Z}[v^{\pm 1}]$-submodule of H generated by $[h, h']_\sigma = hh' - h'\sigma(h)$. By He and Nie [2014], the σ-twisted cocenter $\overline{H} = H/[H, H]_\sigma$ has a standard basis given by $\{T_{\mathcal{O}}\}$, where $\mathcal{O}$ runs over all the σ-conjugacy classes of $\tilde{W}$. Thus for any $w \in \tilde{W}$, we have

$$T_w \equiv \sum_{\mathcal{O}} f_{w,\mathcal{O}} T_{\mathcal{O}} \mod [H, H]_\sigma.$$

The coefficients $f_{w,\mathcal{O}} \in \mathbb{N}[v - v^{-1}]$, which we call the class polynomials (over $v - v^{-1}$). We have the following "dimension=degree" theorem established in He [2014].

Theorem 2.2. *Let $b \in G$ and $w \in \tilde{W}$. Then*

$$\dim(X_w(b)) = \max_{\mathcal{O}; \Psi(\mathcal{O}) = [b]} \frac{1}{2}\big(\ell(w) + \ell(\mathcal{O}) + \deg(f_{w,\mathcal{O}})\big) - \langle \nu_b, 2\rho \rangle.$$

Here $\ell(\mathcal{O})$ is the length of any minimal length element in $\mathcal{O}$ and ρ is the half sum of positive roots in G. Here we use the convention that the dimension of an empty variety and the degree of a zero polynomial are both $-\infty$. Thus the above theorem reduces the nonemptiness question and the dimension formula of $X_w(b)$ to some questions on the class polynomials $f_{w,\mathcal{O}}$ for $\Psi(\mathcal{O}) = [b]$.

The explicit computation of the class polynomials is very difficult at present. Note that there is a close relation between the cocenter and representations of affine Hecke algebras

Ciubotaru and He [2017]. One may hope that some progress in the representation theory of affine Hecke algebras would also advance our knowledge on affine Deligne-Lusztig varieties. At present, we combine the "dimension=degree" theorem together with some Lie-theoretic techniques, and the results on $X_\lambda(b)$ in the affine Grassmannian established previously by arithmetic-geometric method, to obtain some explicit answers to certain questions on $X_w(b)$ and on $X(\mu, b)$.

3 Nonemptiness pattern

3.1 Mazur's inequality. In this subsection, we discuss the non-emptiness patterns of affine Deligne-Lusztig varieties. Here Mazur's inequality plays a crucial role.

In Mazur [1973], Mazur proved that the Hodge slope of any F-crystal is always larger than or equal to the Newton slope of associated isocrystal. The converse was obtained by Kottwitz and Rapoport in R. Kottwitz and Rapoport [2003]. Here we regard the Newton slope and Hodge slope as elements in $\mathbb{Q}_+^n = \{a_1, \cdots, a_n; a_1 \geqslant \cdots \geqslant a_n\}$ and the partial order in $\mathbb{Q}_+^n$ is the dominance order, i.e. $(a_1, \cdots, a_n) \preceq (b_1, \cdots, b_n)$ if and only if $a_1 \leqslant b_1, a_1+a_2 \leqslant b_1+b_2, \cdots, a_1+\cdots+a_{n-1} \leqslant b_1+\cdots+b_{n-1}, a_1+\cdots+a_n = b_1+\cdots+b_n$.

Note that $\mathbb{Q}_+^n$ is the set of rational dominant coweights for GL_n. The dominant order can be defined for the set of rational dominant coweights for any reductive group. This is what we use to describe the nonemptiness pattern of some affine Deligne-Lusztig varieties.

3.2 In the affine Grassmannian. For $X_\lambda(b)$ in the affine Grassmannian, we have a complete answer to the nonemptiness question.

Theorem 3.1. *Let λ be a dominant coweight and $b \in G$. Then $X_\lambda(b) \neq \varnothing$ if and only if $\kappa([b]) = \kappa(\lambda)$ and $\nu_b \preceq \lambda$.*

The "only if" part was proved by Rapoport and Richartz in Rapoport and Richartz [1996], and by Kottwitz in R. E. Kottwitz [2003]. The "if" part was proved by Gashi [2010]. The result also holds if the hyperspecial subgroup of an unramified group is replaced by a maximal special parahoric subgroup of an arbitrary reductive group. This was obtained in He [2014] using the "dimension=degree" Theorem 2.2.

3.3 In the affine flag. Now we consider the variety $X_w(b)$ in the affine flag variety.

(i) We first discuss the case where $[b]$ is basic, i.e., the corresponding Newton point ν_b is central in G (and thus Mazur's inequality is automatically satisfied).

Theorem 3.2. *Let G be a quasi-split group. Let $[b] \in B(G)$ be basic and $w \in \tilde{W}$. Then $X_w(b) \neq \varnothing$ if and only if there is no "Levi obstruction".*

The "Levi obstruction" is defined in terms of the P-alcove elements, introduced by Görtz, Haines, Kottwitz, and Reuman in Görtz, Haines, R. E. Kottwitz, and Reuman [2010]. The explicit definition is technical and we omit it here. This result was conjectured by Görtz, Haines, R. E. Kottwitz, and Reuman [ibid.] for split groups and was established in joint work with Görtz and Nie 2015 for any quasi-split group. Note that the "quasi-split" assumption here is not essential as one may relate $X_w(b)$ for any reductive group G to another affine Deligne-Lusztig variety for the quasi-split inner form of G. We refer to He [2016a, Theorem 2.27] for the explicit statement in the general setting.

(ii) For any nonbasic σ-conjugacy class $[b]$, one may ask for analogues of "Mazur's inequality" and/or the "Levi obstruction" in order to describe the nonemptiness pattern of $X_w(b)$. This is one of the major open problems in this area. We refer to Görtz, Haines, R. E. Kottwitz, and Reuman [2010, Remark 12.1.3] for some discussion in this direction. As a first step, one may consider the conjecture of Görtz-Haines-Kottwitz-Reumann Görtz, Haines, R. E. Kottwitz, and Reuman [ibid., Conjecture 9.5.1 (b)] on the asymptotic behavior of $X_w(b)$ for nonbasic $[b]$. Some affirmative answer to this conjecturewas given in He [2016a, Theorem 2.28] in the case where $[b] = [\epsilon^\lambda]$ for some dominant coweight λ.

3.4 Kottwitz-Rapoport conjecture. To describe the nonemptiness pattern on the union $X(\mu, b)$ of affine Deligne-Lusztig varieties in the affine flag variety, we recall the definition of neutrally acceptable σ-conjugacy classes introduced by Kottwitz in R. E. Kottwitz [1997],

$$B(G, \mu) = \{[b] \in B(G); \kappa([b]) = \kappa(\mu), \nu_b \leqslant \mu^\diamond\},$$

where $\mu^\diamond$ is the Galois average of μ.

By Theorem 3.1, $X_\mu(b) \neq \varnothing$ if and only if $[b] \in B(G, \mu)$. We have a similar result for the union $X(\mu, b)$ of affine Deligne-Lusztig varieties in the affine flag variety.

Theorem 3.3. *Let $[b] \in B(G)$. Then $X(\mu, b) \neq \varnothing$ if and only if $[b] \in B(G, \mu)$.*

This result was conjectured by Kottwitz and Rapoport in R. Kottwitz and Rapoport [2003] and Rapoport [2005]. The "only if" part is a group-theoretic version of Mazur's inequality and was proved by Rapoport and Richartz for unramified groups in Rapoport and Richartz [1996, Theorem 4.2]. The "if" part is the "converse to Mazur's inequality" and was proved by Wintenberger in Wintenberger [2005] for quasi-split groups. The general case in both directions was established in He [2016b] by a different approach, via a detailed analysis of the map $\Psi : B(\tilde{W}) \to B(G)$, of the partial orders on $B(G)$ (an analogy of Grothendieck's conjecture for the loop groups) and of the maximal elements in $B(G, \mu)$ He and Nie [2018].

As we mentioned in Section 3.3, for a single affine Deligne-Lusztig variety $X_w(b)$, one may reduce the case of a general group to the quasi-split case. However, for the union of affine Deligne-Lusztig varieties, the situation is different. There is no relation between the admissible set $\mathrm{Adm}(\mu)$ (and hence $X(\mu, b)$) for an arbitrary reductive group and its quasi-split inner form. This adds essential difficulties in the study of $X(\mu, b)$ for non quasi-split groups.

Rad and Hartl in Rad and Hartl [2016] established the analogue of the Langlands-Rapoport conjecture Langlands and Rapoport [1987] for the rational points in the moduli stacks of global G-shtukas, for arbitrary connected reductive groups and arbitrary parahoric level structure. They described the rational points as a disjoint union over isogeny classes of global G-Shtukas, and then used Theorem 3.3 to determine which isogeny classes are nonempty.

4 Dimension formula

4.1 In the affine Grassmannian. For $X_\lambda(b)$ in the affine Grassmannian, we have an explicit dimension formula.

Theorem 4.1. *Let λ be a dominant coweight and $b \in G$. If $X_\lambda(b) \neq \varnothing$, then*

$$\dim X_\lambda(b) = \langle \lambda - \nu_b, \rho \rangle - \frac{1}{2} \mathit{def}_G(b),$$

where $\mathit{def}_G(b)$ is the defect of b.

The dimension formula of $X_\lambda(b)$ was conjectured by Rapoport in Rapoport [2005], inspired by Chai's work Chai [2000]. The current reformulation is due to R. E. Kottwitz [2006]. For split groups, the conjectural formula was obtained by Görtz, Haines, R. E. Kottwitz, and Reuman [2006] and Viehmann [2006]. The conjectural formula for general quasi-split unramified groups was obtained independently by X. Zhu [2017] and Hamacher [2015a].

4.2 In the affine flag variety. Now we consider $X_w(b)$ in the affine flag variety.

Theorem 4.2. *Let $[b] \in B(G)$ be basic and $w \in \tilde{W}$ be an element in the shrunken Weyl chamber (i.e., the lowest two-sided cell of $\tilde{W}$). If $X_w(b) \neq \varnothing$, then*

$$\dim X_w(b) = \frac{1}{2}(\ell(w) + \ell(\eta_\sigma(w)) - \mathit{def}_G(b)).$$

Here $\eta_\sigma : \tilde{W} \to W_0$ is defined in Görtz, Haines, R. E. Kottwitz, and Reuman [2010].

This dimension formula was conjectured by Görtz, Haines, Kottwitz, and Reuman in Görtz, Haines, R. E. Kottwitz, and Reuman [ibid.] for split groups and was established for residually split groups in He [2014]. The proof in He [ibid.] is based on the "dimension=degree" Theorem 2.2, some results on the σ-twisted cocenter $\overline{H}$ of affine Hecke algebra H, together with the dimension formula of $X_\lambda(b)$ (which was only known for split groups at that time). The dimension formula for arbitrary reductive groups (under the same assumption on b and w) is obtained by the same argument in He [ibid.], once the dimension formula of $X_\lambda(b)$ for quasi-split unramified groups became available, cf. Theorem 4.1.

Note that the assumption that w is contained in the lowest two-sided cell is an essential assumption here. A major open problem is to understand the dimension of $X_w(b)$ for $[b]$ basic, when w is in the critical stripes (i.e., outside the lowest two-sided cell). So far, no conjectural dimension formula has been formulated. However, the "dimension=degree" Theorem 2.2 and the explicit computation in low rank cases Görtz, Haines, R. E. Kottwitz, and Reuman [2010] indicate that this problem might be closely related to the theory of Kazhdan-Lusztig cells. I expect that further progress on the affine cellularity of affine Hecke algebras, which is a big open problem in representation theory, might shed new light on the study of $\dim X_w(b)$.

I also would like to point out that affine Deligne-Lusztig varieties in affine Grassmannians are equi-dimensional, while in general affine Deligne-Lusztig varieties in the affine flag varieties are not equi-dimensional.

4.3 Certain unions. We will see in Section 8 that for certain pairs (G, μ), $X(\mu, b)$ admits some simple geometric structure. In these cases, one may write down an explicit dimension formula for $X(\mu, b)$. Outide these case, very little is known for $\dim X(\mu, b)$.

Here we mention one difficult case: the Siegel modular variety case. Here $G = Sp_{2g}$ and μ is the minuscule coweight. It was studied by Görtz and Yu in Görtz and Yu [2010], in which they showed that for basic $[b]$, $\dim X(\mu, b) = \frac{g^2}{2}$ if g is even and $\frac{g(g-1)}{2} \leqslant \dim X(\mu, b) \leqslant [\frac{g^2}{2}]$ if g is odd. It would be interesting to determine the exact dimension when g is odd.

5 Hodge–Newton decomposition

To study the set-theoretic and geometric properties of affine Deligne-Lusztig varieties, a very useful tool is to reduce the study of affine Deligne-Lusztig varieties of a connected reductive group to certain affine Deligne-Lusztig varieties of its Levi subgroups. Such reduction is achieved by the Hodge-Newton decomposition, which originated in Katz's work Katz [1979] on F-crystals with additional structures. In this section, we discuss its variation for affine Deligne-Lusztig varieties in affine Grassmannians, and further development on affine Deligne-Lusztig varieties in affine flag varieties, and on the union of affine Deligne-Lusztig varieties.

5.1 In the affine Grassmannian. For affine Deligne-Lusztig varieties in the affine Grassmannian, Kottwitz in R. E. Kottwitz [2003] (see also Viehmann [2008]) established the following Hodge-Newton decomposition, which is the group-theoretic generalization of Katz's result. Here the pair (λ, b) is called Hodge-Newton decomposable with respect to a proper Levi subgroup M if $b \in M$ and λ and b have the same image under the Kottwitz's map κ_M for M.

Theorem 5.1. *Let M be a Levi subgroup of G and (λ, b) be Hodge-Newton decomposable with respect to M. Then the natural map $X^M_\lambda(b) \to X^G_\lambda(b)$ is an isomorphism.*

5.2 In the affine flag variety. For affine Deligne-Lusztig varieties in affine flag varieties, the situation is more complicated, as the Hodge-Newton decomposability condition on the pairs (w, b) is rather difficult. As pointed out in Görtz, Haines, R. E. Kottwitz, and Reuman [2010], "It is striking that the notion of P-alcove, discovered in the attempt to understand the entire emptiness pattern for the $X_x(b)$ when b is basic, is also precisely the notion needed for our Hodge-Newton decomposition."

The Hodge-Newton decomposition for $X_w(b)$ was established by Görtz, Haines, Kottwitz and Reuman in Görtz, Haines, R. E. Kottwitz, and Reuman [ibid.].

Theorem 5.2. *Suppose that $P = MN$ is a semistandard Levi subgroup of G and $w \in \tilde{W}$ is a P-alcove element in the sense of Görtz, Haines, R. E. Kottwitz, and Reuman [ibid.]. Let $b \in M$. Then the natural map $X^M_w(b) \to X^G_w(b)$ induces a bijection*

$$J^M_b \backslash X^M_w(b) \cong J^G_b \backslash X^G_w(b).$$

5.3 Certain unions. For $X(\mu, b)$, the Hodge-Newton decomposability condition is still defined on the pair (μ, b). However, the precise condition is more complicated than in Section 5.1 as we consider arbitrary connected reductive groups, not only the unramified ones. We refer to Goertz, He, and Nie [2016, Definition 2.1] for the precise definition. The following Hodge-Newton decomposition for $X(\mu, b)$ was established in a joint work with Görtz and Nie 2016.

Theorem 5.3. *Suppose that (μ, b) is Hodge-Newton decomposable with respect to some proper Levi subgroup. Then*

$$X(\mu, b) \cong \bigsqcup_{P'=M'N'} X^{M'}(\mu_{P'}, b_{P'}),$$

where P' runs through a certain finite set of semistandard parabolic subgroups. The subsets in the union are open and closed.

We refer to Goertz, He, and Nie [2016, Theorem 3.16] for the precise statement. Note that an essential new feature is that unlike the Hodge-Newton decomposition of a single affine Deligne-Lusztig variety (e.g. $X_\lambda(b)$ or $X_w(b)$) where only one Levi subgroup is involved, in the Hodge-Newton decomposition of $X(\mu, b)$ several Levi subgroups are involved.

Thus, the statement here is more complicated than the Hodge-Newton decomposition of $X_\lambda(b)$ and $X_w(b)$. But this is consistent with the fact that the Newton strata in the special fiber of Shimura varieties with Iwahori level structure are more complicated than those with hyperspecial level structure. I believe that the Hodge-Newton decomposition here would help us to overcome some of the difficulties occurring in the study of Shimura varieties with Iwahori level structure (as well as arbitrary parahoric level structures). We will see some results in this direction in Section 6 and in Section 8.

6 Connected components

In this subsection, we discuss the set of connected components of some closed affine Deligne-Lusztig varieties, e.g.

$$X_{\preceq\lambda}(b) := \cup_{\lambda' \preceq \lambda} X_{\lambda'}(b) \text{ and } X(\mu, b) = \cup_{w \in \text{Adm}(\mu)} X_w(b).$$

The explicit description of the set of connected components has some important applications in number theory, which we will mention later.

Note that affine Grassmannians and affine flag varieties are not connected in general, and their connected components are indexed by $\pi_1(G)_{\Gamma_0}$. This gives the first obstruction to the connectedness. The second obstruction comes from the Hodge-Newton decomposition, which we discussed in Section 5. One may expect that these are the only obstructions. We have the following results.

Theorem 6.1. *Assume that G is an unramified simple group and that (λ, b) is Hodge-Newton indecomposable. Then*

$$\pi_0(X_{\preceq\lambda}(b)) \cong \pi_1(G)^{\sigma}_{\Gamma_0}.$$

This was first proved by Viehmann for split groups, and then by M. Chen, Kisin, and Viehmann [2015] for quasi-split unramified groups and for λ minuscule. The description of $\pi_0(X_{\preceq\lambda}(b))$ for G quasi-split unramified, and λ non-minuscule, was conjectured in M. Chen, Kisin, and Viehmann [ibid.] and was established by Nie [2015].

Note that the minuscule coweight case is especially important for applications in number theory. Kisin [2017] proved the Langlands-Rapoport conjecture for mod-p points on Shimura varieties of *abelian type* with *hyperspecial level structure*. Compared to the function field analogous of Langlands-Rapoport conjecture Rad and Hartl [2016], there are extra complication coming from algebraic geometry and the explicit description of the connected components of $X(\mu, b)$ in M. Chen, Kisin, and Viehmann [2015] is used in an essential way to overcome the complication.

Theorem 6.2. *Let μ be a dominant coweight and $b \in G$. Assume that $[b] \in B(G, \mu)$ and that (μ, b) is Hodge-Newton indecomposable. Then*

(1) If $[b]$ is basic, then $\pi_0(X(\mu, b)) \cong \pi_1(G)^{\sigma}_{\Gamma_0}$.

(2) If G is split, then $\pi_0(X(\mu, b)) \cong \pi_1(G)$.

Here part (1) was obtained in joint work with Zhou 2016. As an application, we verified the Axioms in He and Rapoport [2017] for certain PEL type Shimura varieties. In He and Zhou [2016], the set of connected components of $X(\mu, b)$ was also studied for nonbasic b. We proved the in a residually split group, the set of connected components is "controlled" by the set of straight elements, together with the obstruction from the corresponding Levi subgroup. Combined with the work of Zhou [2017], we verified in the residually split case, the description of the mod-p isogeny classes on Shimura varieties conjectured by Langlands and Rapoport [1987]. Part (2) is recent work of L. Chen and Nie [2017].

We would like to point out that in the statement, the following two conditions are essential:

- The σ-conjugacy class $[b]$ is neutrally acceptable, i.e. $[b] \in B(G, \mu)$. This condition comes from the Kottwitz-Rapoport conjecture (see Theorem 3.3).
- The pair (μ, b) is Hodge-Newton indecomposable. In the general case, we need to apply the Hodge-Newton decomposition (see Theorem 5.3). As a consequence, several $\pi_1(M)$ are involved in the description of $\pi_0(X(\mu, b))$ in general.

7 Arbitrary parahoric level structure

7.1 Parahoric level versus Iwahori level. Let $K' \supset I$ be a standard parahoric subgroup of G and $W_{K'}$ be the finite Weyl group of K'. We define

$$X(\mu, b)_{K'} = \{gK' \in G/K'; g^{-1} b \sigma(g) \in K' \operatorname{Adm}(\mu) K'\}.$$

If $K' = I$, then $X(\mu, b)_{K'} = X(\mu, b)$. If G is unramified, μ is minuscule and $K' = K$ is a hyperspecial parahoric subgroup, then $X(\mu, b)_{K'} = X_\mu(b)$. As we have mentioned, the varieties $X_\mu(b)$ (resp. $X(\mu, b)$) serve as group-theoretic models for the Newton strata in the special fiber of Shimura varieties with hyperspecial (resp. Iwahori) level structure. The variety $X(\mu, b)_{K'}$ plays the same role in the study of Shimura varieties with arbitrary parahoric level structure.

The following result relates $X(\mu, b)_{K'}$ for an arbitrary parahoric subgroup K' with $X(\mu, b)$ (for the Iwahori subgroup I).

Theorem 7.1. *The projection map $G/I \to G/K'$ induces a surjection*

$$X(\mu, b) \twoheadrightarrow X(\mu, b)_{K'}.$$

This was conjectured by Kottwitz and Rapoport in Kudla and Rapoport [2011] and Rapoport [2005] and was proved in He [2016b]. This fact allows one to reduce many questions (e.g. nonemptiness pattern, connected components, etc.) of $X(\mu, b)_{K'}$ for arbitrary K' to the same questions for $X(\mu, b)$. In fact, the statements in Theorem 3.3 and Theorem 6.2 hold if $X(\mu, b)$ is replaced by $X(\mu, b)_{K'}$ for an arbitrary parahoric subgroup K'.

7.2 Lusztig's G-stable pieces. I would like to draw attention to some crucial ingredient in the proof, which has important applications in arithmetic geometry.

Note that $I \operatorname{Adm}(\mu) I \subsetneq K' \operatorname{Adm}(\mu) K'$ if $I \subsetneq K'$. In order to show that $X(\mu, b) \to X(\mu, b)_{K'}$ is surjective, one needs to have some decomposition of $K' \operatorname{Adm}(\mu) K'$, finer than the decomposition into K' double cosets. The idea of the sought-after decomposition is essentially due to Lusztig. In 2004, Lusztig introduced G-stable pieces for reductive groups over algebraically closed fields. The closure relation between G-stable pieces was determined in He [2007b] and a more systematic approach using the " partial conjugation action" technique was given later in He [2007a]. The notion and the closure relation of G-stable pieces also found application in arithmetic geometry, e.g. in the work of Pink, Wedhorn and Ziegler on algebraic zip data Pink, Wedhorn, and Ziegler [2011].

7.3 Ekedahl-Kottwitz-Oort-Rapoport stratification. Lusztig's idea was also adapted to loop groups, first with a hyperspecial parahoric subgroup, independently by the author 2011, and by Viehmann [2014]. It was used it to define the Ekedahl-Oort stratification of a general Shimura variety.

The desired decomposition of $K' \operatorname{Adm}(\mu) K'$ for an arbitrary parahoric subgroup K' was given in He [2016b] as

$$K' \operatorname{Adm}(\mu) K' = \sqcup_{w \in {}^{K'}\tilde{W} \cap \operatorname{Adm}(\mu)} K' \cdot_\sigma IwI,$$

where ${}^{K'}\tilde{W}$ is the set of minimal length elements in $W_{K'} \backslash \tilde{W}$ and $\cdot_\sigma$ means the σ-conjugation action. This decomposition is used in joint work with Rapoport 2017 to define the Ekedahl-Kottwitz-Oort-Rapoport stratification of Shimura varieties with arbitrary parahoric level structure. This stratification interpolates between the Kottwitz-Rapoport stratification in the case of the Iwahori level structure and the Ekedahl-Oort stratification Viehmann [2014] in the case of hyperspecial level structure.

8 Affine Deligne-Lusztig varieties with simple geometric structure

8.1 Simple geometric structure for some $X(\mu, b_0)_{K'}$. The geometric structure of $X(\mu, b_0)_{K'}$ for basic b_0 is rather complicated in general. However, in certain cases, $X(\mu, b_0)_{K'}$ admit a simple description. The first nontrivial example is due to Vollaard and Wedhorn in Vollaard and Wedhorn [2011]. They showed that $X_\mu(b_0)$ for an unramified unitary group of signature $(1, n-1)$ and $\mu = (1, 0, \cdots, 0)$ (and for hyperspecial parahoric level structure), is a union of classical Deligne-Lusztig varieties, and the index set and the closure relations between the strata are encoded in a Bruhat-Tits building. Since then, this question has attracted significant attention. We mention the work of Rapoport, Terstiege, and Wilson [2014] on ramified unitary groups, of Howard and Pappas [2014], Howard and Pappas [2017] on orthogonal groups, of Tian and Xiao [2016] in the Hilbert-Blumenthal case. In all these works, the parahoric subgroups involved are hyperspecial parahoric subgroups or certain maximal parahoric subgroups. The analogous group-theoretic question for maximal parahoric subgroups was studied in joint work with Görtz Görtz and He [2015].

Note that these simple descriptions of closed affine Deligne-Lusztig varieties (and the corresponding basic locus of Shimura varieties) have been used, with great success, towards applications in number theory: to compute intersection numbers of special cycles, as in the Kudla-Rapoport program Kudla and Rapoport [2011] or in work Rapoport, Terstiege, and W. Zhang [2013], Li and Y. Zhu [2017] towards Zhang's Arithmetic Fundamental Lemma W. Zhang [2012]; and to prove the Tate conjecture for certain Shimura varieties Tian and Xiao [2014], Helm, Tian, and Xiao [2017].

The work of Vollaard and Wedhorn [2011], Rapoport, Terstiege, and Wilson [2014], Howard and Pappas [2014], Howard and Pappas [2017], Tian and Xiao [2016] focused on specific Shimura varieties with certain maximal parahoric level structure. The work Görtz and He [2015] studied the analogous group-theoretic question for arbitrary reductive groups. The conceptual interpretation on the occurrence of classical Deligne-Lusztig varieties was given; however, a large part of the work in Görtz and He [ibid.] was still obtained by brute force.

8.2 Some equivalent conditions. From the Lie-theoretic point of view, one would like to consider not only the maximal parahoric subgroups, but all parahoric subgroups; and one would like to have a conceptual understanding on the following question:

When and why is $X(\mu, b_0)_{K'}$ naturally a union of classical Deligne-Lusztig varieties? This was finally achieved in joint work with Görtz and Nie 2015 as follows

Theorem 8.1. *Assume that G is simple, μ is a dominant coweight of G and K' is a parahoric subgroup. Then the following conditions are equivalent:*

- *For basic $[b_0] \in B(G, \mu)$, $X(\mu, b_0)_{K'}$ is naturally a union of classical Deligne-Lusztig varieties;*
- *For any nonbasic $[b] \in B(G, \mu)$, $\dim X(\mu, b)_{K'} = 0$;*
- *The pair (μ, b) is Hodge-Newton decomposable for any nonbasic $[b] \in B(G, \mu)$;*
- *The coweight μ is minute for G.*

Here the minute condition is an explicit combinatorial condition on the coweight μ. For quasi-split groups, it means that for any σ-orbit $\mathcal{O}$ on the set of simple roots, we have $\sum_{i \in \mathcal{O}} \langle \mu, \omega_i \rangle \leqslant 1$. For non quasi-split groups, the condition is more involved and we refer to Goertz, He, and Nie [2016, Definition 2.2] for the precise definition. It is also worth mentioning that it is not very difficult to classify the pairs (G, μ) with the minute condition. In Goertz, He, and Nie [ibid., Theorem 2.5], a complete list of the cases is obtained, where $X(\mu, b_0)_{K'}$ is naturally a union of classical Deligne-Lusztig varieties.

Fargues and Rapoport conjectured that for p-adic period domains, the weakly admissible locus coincides with the admissible locus if and only if the pair (μ, b) is Hodge-Newton decomposable for any nonbasic $[b] \in B(G, \mu)$ (cf. Goertz, He, and Nie [ibid., Conjecture 0.1]). This conjectured is established in a very recent preprint M. Chen, Fargues, and Shen [2017] by Chen, Fargues and Shen.

8.3 Further remarks. From the Lie-theoretic point of view, there are some quite striking new features in Theorem 8.1:

1. The relations between the variety $X(\mu, b)_{K'}$ for the basic σ-conjugacy class and for nonbasic σ-conjugacy classes;
2. The relation between the condition that $X(\mu, b_0)_{K'}$ has a simple description and the Hodge-Newton decomposability condition;
3. The existence of a simple description of $X(\mu, b_0)_{K'}$ is independent of the parahoric subgroup K'.

Note that part (1) and part (2) are new even for the specific Shimura varieties with hyperspecial level structure considered in the previous works. Part (3) is the most mysterious one. In Goertz, He, and Nie [2016], we state that "We do not see any reason why this independence of the parahoric could be expected a priori, but it is an interesting parallel with the question when the weakly admissible and admissible loci in the rigid analytic period domain coincide."

For applications to number theory, one needs to consider the fraction field of the Witt ring instead of the formal Laurent series field $\overline{\mathbb{F}}_q((\epsilon))$. In that setting, we have a similar, but weaker result, namely, $X(\mu, b_0)_{K'}$ is naturally a union of classical Deligne-Lusztig varieties as perfect schemes. It is expected that the structural results hold without perfection, as indicated in the special cases established in the papers mentioned in Section 8.1.

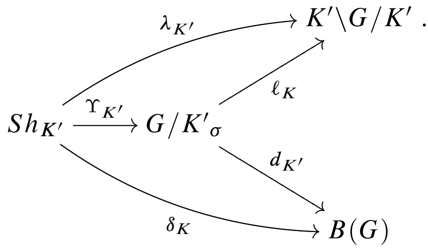
9 Some applications to Shimura varieties

In the last subsection, we give a very brief discussion of some applications to arithmetic geometry.

9.1 Some characteristic subsets. The study of some characteristic subsets in the special fiber of a Shimura variety is a central topic in arithmetic geometry. We mention the *Newton strata*, the *Ekedahl-Oort strata* for the hyperspecial level structure and the *Kottwitz-Rapoport strata* for the Iwahori level structure. Concerning these stratifications, there are many interesting questions one may ask, e.g. which strata are nonempty, what is the relation between these various stratifications, etc.. These questions have been intensively studied in recent years and there is a large body of literature on these questions. Among them, we mention the work of Viehmann and Wedhorn [2013] on the nonemptiness of Newton strata and Ekedahl-Oort strata for PEL type Shimura varieties with hyperspecial level structure, the work of M. Kisin, Madapusi, and Shin [n.d.] on the nonemptiness of the basic Newton stratum, the work of Hamacher [2015b] on the closure relation between Newton strata, and the work of Wedhorn [1999] and Moonen [2004] on the density of the μ-ordinary locus (i.e. the Newton stratum corresponding to ϵ^{μ}). We refer to He and Rapoport [2017, Introduction] and Viehmann [2015] for more references.

9.2 An axiomatic approach. In the works mentioned above, both algebro-geometric and Lie-theoretic methods are involved, and are often mixed together.

In joint work with He and Rapoport [2017], we purposed an axiomatic approach to the study of these characteristic subsets in a general Shimura variety. We formulated five axioms, based on the existence of integral models of Shimura varieties (which have been established in various cases by the work of Rapoport and Zink [1996], Kisin and Pappas [2015]), the existence of the following commutative diagram and some compatibility conditions:



Here K' is a parahoric subgroup, $Sh_{K'}$ is the special fiber of a Shimura variety with K' level structure, and G/K'_σ is the set-theoretic quotient of G by the σ-conjugation action of K'.

As explained in Goertz, He, and Nie [2016, §6.2], affine Deligne-Lusztig varieties are involved in the diagram in an essential way, via the bijection

$$J_b \backslash X(\mu, b)_{K'} \xrightarrow{\sim} d_{K'}^{-1}([b]) \cap \ell_{K'}^{-1}(K' \operatorname{Adm}(\{\mu\})K').$$

9.3 Some applications and current status of the axioms. It is shown in He and Rapoport [2017] that under those axioms, the Newton strata, the Ekedahl-Oort strata, the Kottwitz-Rapoport strata, and the Ekedahl-Kottwitz-Oort-Rapoport strata discussed in Section 7, are all nonempty in their natural range. Furthermore, under those axioms several relations between these various stratifications are also established in He and Rapoport [ibid.].

Following He and Rapoport [ibid.], Shen and C. Zhang [2017] studied the geometry of good reductions of Shimura varieties of abelian type. They established basic properties of these characteristic subsets, including nonemptiness, closure relations and dimension formula and some relations between these stratifications.

In joint work with Nie 2017, based on the framework of He and Rapoport [2017], we studied the density problem of the μ-ordinary locus. Under the axioms of He and Rapoport [ibid.] we gave several explicit criteria on the density of the μ-ordinary locus.

Algebraic geometry is essential in the verification of these axioms. For PEL type Shimura varieties associated to unramified groups of type A and C and to odd ramified unitary groups, the axioms are verified in joint work with Zhou 2016. For Shimura varieties of Hodge type, most of the axioms are verified in recent work of Zhou [2017].

References

Bhargav Bhatt and Peter Scholze (2017). "Projectivity of the Witt vector affine Grassmannian". *Invent. Math.* 209.2, pp. 329–423. MR: 3674218 (cit. on p. 1364).

F. Bruhat and J. Tits (1972). "Groupes réductifs sur un corps local". *Inst. Hautes Études Sci. Publ. Math.* 41, pp. 5–251. MR: 0327923 (cit. on p. 1365).

Ching-Li Chai (2000). "Newton polygons as lattice points". *Amer. J. Math.* 122.5, pp. 967–990. MR: 1781927 (cit. on p. 1371).

Ling Chen and Sian Nie (Mar. 2017). "Connected components of closed affine Deligne-Lusztig varieties". arXiv: 1703.02476 (cit. on p. 1375).

M. Chen, L. Fargues, and X. Shen (Oct. 2017). "On the structure of some p-adic period domains". arXiv: 1710.06935 (cit. on p. 1378).

Miaofen Chen, Mark Kisin, and Eva Viehmann (2015). "Connected components of affine Deligne-Lusztig varieties in mixed characteristic". *Compos. Math.* 151.9, pp. 1697–1762. MR: 3406443 (cit. on pp. 1374, 1375).

Dan Ciubotaru and Xuhua He (2017). "Cocenters and representations of affine Hecke algebras". *J. Eur. Math. Soc. (JEMS)* 19.10, pp. 3143–3177. MR: 3713003 (cit. on p. 1369).

P. Deligne and G. Lusztig (1976). "Representations of reductive groups over finite fields". *Ann. of Math. (2)* 103.1, pp. 103–161. MR: 0393266 (cit. on pp. 1363, 1364, 1368).

D. Gaitsgory (2001). "Construction of central elements in the affine Hecke algebra via nearby cycles". *Invent. Math.* 144.2, pp. 253–280. MR: 1826370 (cit. on p. 1366).

Qëndrim R. Gashi (2010). "On a conjecture of Kottwitz and Rapoport". *Ann. Sci. Éc. Norm. Supér. (4)* 43.6, pp. 1017–1038. MR: 2778454 (cit. on p. 1369).

Meinolf Geck and Götz Pfeiffer (1993). "On the irreducible characters of Hecke algebras". *Adv. Math.* 102.1, pp. 79–94. MR: 1250466.

Ulrich Goertz, Xuhua He, and Sian Nie (Oct. 2016). "Fully Hodge-Newton decomposable Shimura varieties". arXiv: 1610.05381 (cit. on pp. 1373, 1374, 1378–1380).

Ulrich Görtz, Thomas J. Haines, Robert E. Kottwitz, and Daniel C. Reuman (2006). "Dimensions of some affine Deligne-Lusztig varieties". *Ann. Sci. École Norm. Sup. (4)* 39.3, pp. 467–511. MR: 2265676 (cit. on p. 1371).

– (2010). "Affine Deligne-Lusztig varieties in affine flag varieties". *Compos. Math.* 146.5, pp. 1339–1382. MR: 2684303 (cit. on pp. 1367, 1370–1373).

Ulrich Görtz and Xuhua He (2015). "Basic loci of Coxeter type in Shimura varieties". *Camb. J. Math.* 3.3, pp. 323–353. MR: 3393024 (cit. on p. 1377).

Ulrich Görtz, Xuhua He, and Sian Nie (2015). "**P**-alcoves and nonemptiness of affine Deligne-Lusztig varieties". *Ann. Sci. Éc. Norm. Supér. (4)* 48.3, pp. 647–665. MR: 3377055 (cit. on pp. 1370, 1378).

Ulrich Görtz and Chia-Fu Yu (2010). "Supersingular Kottwitz-Rapoport strata and Deligne-Lusztig varieties". *J. Inst. Math. Jussieu* 9.2, pp. 357–390. MR: 2602029 (cit. on p. 1372).

Thomas J. Haines and Ngô Bao Châu (2002). "Alcoves associated to special fibers of local models". *Amer. J. Math.* 124.6, pp. 1125–1152. MR: 1939783 (cit. on p. 1366).

Thomas J. Haines and Xuhua He (2017). "Vertexwise criteria for admissibility of alcoves". *Amer. J. Math.* 139.3, pp. 769–784. MR: 3650232 (cit. on p. 1366).

Paul Hamacher (2015a). "The dimension of affine Deligne-Lusztig varieties in the affine Grassmannian". *Int. Math. Res. Not. IMRN* 23, pp. 12804–12839. MR: 3431637 (cit. on p. 1371).

– (2015b). "The geometry of Newton strata in the reduction modulo p of Shimura varieties of PEL type". *Duke Math. J.* 164.15, pp. 2809–2895. MR: 3430453 (cit. on p. 1379).

X. He and M. Rapoport (2017). "Stratifications in the reduction of Shimura varieties". *Manuscripta Math.* 152.3-4, pp. 317–343. MR: 3608295 (cit. on pp. 1375, 1376, 1379, 1380).

Xuhua He (2007a). "Minimal length elements in some double cosets of Coxeter groups". *Adv. Math.* 215.2, pp. 469–503. MR: 2355597 (cit. on p. 1376).

– (2007b). "The G-stable pieces of the wonderful compactification". *Trans. Amer. Math. Soc.* 359.7, pp. 3005–3024. MR: 2299444 (cit. on p. 1376).

– (2011). "Closure of Steinberg fibers and affine Deligne-Lusztig varieties". *Int. Math. Res. Not. IMRN* 14, pp. 3237–3260. MR: 2817678 (cit. on p. 1376).

– (2014). "Geometric and homological properties of affine Deligne-Lusztig varieties". *Ann. of Math. (2)* 179.1, pp. 367–404. MR: 3126571 (cit. on pp. 1367–1369, 1372).

– (2016a). "Hecke algebras and p-adic groups". In: *Current developments in mathematics 2015*. Int. Press, Somerville, MA, pp. 73–135. MR: 3642544 (cit. on p. 1370).

– (2016b). "Kottwitz-Rapoport conjecture on unions of affine Deligne-Lusztig varieties". *Ann. Sci. Éc. Norm. Supér. (4)* 49.5, pp. 1125–1141. MR: 3581812 (cit. on pp. 1370, 1376).

Xuhua He and Sian Nie (2014). "Minimal length elements of extended affine Weyl groups". *Compos. Math.* 150.11, pp. 1903–1927. MR: 3279261 (cit. on p. 1368).

– (2017). "On the μ-ordinary locus of a Shimura variety". *Adv. Math.* 321, pp. 513–528. MR: 3715719 (cit. on p. 1380).

– (2018). "On the acceptable elements" (cit. on p. 1370).

Xuhua He and Rong Zhou (Oct. 2016). "On the connected components of affine Deligne-Lusztig varieties". arXiv: 1610.06879 (cit. on pp. 1375, 1380).

David Helm, Yichao Tian, and Liang Xiao (2017). "Tate cycles on some unitary Shimura varieties mod p". *Algebra Number Theory* 11.10, pp. 2213–2288. MR: 3744356 (cit. on p. 1377).

Benjamin Howard and Georgios Pappas (2014). "On the supersingular locus of the $\mathrm{GU}(2,2)$ Shimura variety". *Algebra Number Theory* 8.7, pp. 1659–1699. MR: 3272278 (cit. on p. 1377).

– (2017). "Rapoport-Zink spaces for spinor groups". *Compos. Math.* 153.5, pp. 1050–1118. MR: 3705249 (cit. on p. 1377).

N. Iwahori and H. Matsumoto (1965). "On some Bruhat decomposition and the structure of the Hecke rings of $\mathfrak{p}$-adic Chevalley groups". *Inst. Hautes Études Sci. Publ. Math.* 25, pp. 5–48. MR: 0185016 (cit. on p. 1365).

Nicholas M. Katz (1979). "Slope filtration of F-crystals". In: *Journées de Géométrie Algébrique de Rennes (Rennes, 1978), Vol. I*. Vol. 63. Astérisque. Soc. Math. France, Paris, pp. 113–163. MR: 563463 (cit. on p. 1372).

M. Kisin and G. Pappas (Dec. 2015). "Integral models of Shimura varieties with parahoric level structure". To appear in Publ. Math. IHES. arXiv: 1512.01149 (cit. on p. 1379).

Mark Kisin (2017). "mod p points on Shimura varieties of abelian type". *J. Amer. Math. Soc.* 30.3, pp. 819–914. MR: 3630089 (cit. on p. 1375).

R. Kottwitz and M. Rapoport (2000). "Minuscule alcoves for GL_n and $G\mathrm{Sp}_{2n}$". *Manuscripta Math.* 102.4, pp. 403–428. MR: 1785323 (cit. on p. 1366).

– (2003). "On the existence of F-crystals". *Comment. Math. Helv.* 78.1, pp. 153–184. MR: 1966756 (cit. on pp. 1369, 1370).

Robert E. Kottwitz (1985). "Isocrystals with additional structure". *Compositio Math.* 56.2, pp. 201–220. MR: 809866 (cit. on p. 1367).

– (1997). "Isocrystals with additional structure. II". *Compositio Math.* 109.3, pp. 255–339. MR: 1485921 (cit. on pp. 1367, 1370).

– (2003). "On the Hodge-Newton decomposition for split groups". *Int. Math. Res. Not.* 26, pp. 1433–1447. MR: 1976046 (cit. on pp. 1369, 1373).

– (2006). "Dimensions of Newton strata in the adjoint quotient of reductive groups". *Pure Appl. Math. Q.* 2.3, Special Issue: In honor of Robert D. MacPherson. Part 1, pp. 817–836. MR: 2252119 (cit. on p. 1371).

Stephen Kudla and Michael Rapoport (2011). "Special cycles on unitary Shimura varieties I. Unramified local theory". *Invent. Math.* 184.3, pp. 629–682. MR: 2800697 (cit. on pp. 1376, 1377).

R. P. Langlands and M. Rapoport (1987). "Shimuravarietäten und Gerben". *J. Reine Angew. Math.* 378, pp. 113–220. MR: 895287 (cit. on pp. 1371, 1375).

Chao Li and Yihang Zhu (2017). "Remarks on the arithmetic fundamental lemma". *Algebra Number Theory* 11.10, pp. 2425–2445. MR: 3744362 (cit. on pp. 1364, 1377).

G. Lusztig (2004). "Parabolic character sheaves. II". *Mosc. Math. J.* 4.4, pp. 869–896, 981. MR: 2124170 (cit. on p. 1376).

– (June 2010). "Bruhat decomposition and applications". arXiv: 1006.5004 (cit. on p. 1363).

– (2012). "From conjugacy classes in the Weyl group to unipotent classes, III". *Represent. Theory* 16, pp. 450–488. MR: 2968566 (cit. on p. 1364).

George Lusztig (1984). *Characters of reductive groups over a finite field*. Vol. 107. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xxi+384. MR: 742472 (cit. on p. 1364).

M. Kisin, K. Madapusi, and S.-W. Shin (n.d.). "Honda–Tate theory for Shimura varieties of Hodge type" (cit. on p. 1379).

B. Mazur (1973). "Frobenius and the Hodge filtration (estimates)". *Ann. of Math. (2)* 98, pp. 58–95. MR: 0321932 (cit. on p. 1369).

Ben Moonen (2004). "Serre-Tate theory for moduli spaces of PEL type". *Ann. Sci. École Norm. Sup. (4)* 37.2, pp. 223–269. MR: 2061781 (cit. on p. 1379).

Sian Nie (Nov. 2015). "Connected components of closed affine Deligne-Lusztig varieties in affine Grassmannians". arXiv: 1511.04677 (cit. on p. 1374).

G. Pappas and M. Rapoport (2008). "Twisted loop groups and their affine flag varieties". *Adv. Math.* 219.1. With an appendix by T. Haines and Rapoport, pp. 118–198. MR: 2435422 (cit. on pp. 1364, 1366).

Richard Pink, Torsten Wedhorn, and Paul Ziegler (2011). "Algebraic zip data". *Doc. Math.* 16, pp. 253–300. MR: 2804513 (cit. on p. 1376).

Esmail Arasteh Rad and Urs Hartl (May 2016). "Langlands-Rapoport Conjecture Over Function Fields". arXiv: 1605.01575 (cit. on pp. 1371, 1375).

M. Rapoport and M. Richartz (1996). "On the classification and specialization of F-isocrystals with additional structure". *Compositio Math.* 103.2, pp. 153–181. MR: 1411570 (cit. on pp. 1369, 1370).

M. Rapoport and Th. Zink (1996). *Period spaces for p-divisible groups*. Vol. 141. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xxii+324. MR: 1393439 (cit. on p. 1379).

Michael Rapoport (2005). "A guide to the reduction modulo p of Shimura varieties". *Astérisque* 298. Automorphic forms. I, pp. 271–318. MR: 2141705 (cit. on pp. 1365, 1370, 1371, 1376).

Michael Rapoport, Ulrich Terstiege, and Sean Wilson (2014). "The supersingular locus of the Shimura variety for $\mathrm{GU}(1, n-1)$ over a ramified prime". *Math. Z.* 276.3-4, pp. 1165–1188. MR: 3175176 (cit. on p. 1377).

Michael Rapoport, Ulrich Terstiege, and Wei Zhang (2013). "On the arithmetic fundamental lemma in the minuscule case". *Compos. Math.* 149.10, pp. 1631–1666. MR: 3123304 (cit. on pp. 1364, 1377).

Xu Shen and Chao Zhang (July 2017). "Stratifications in good reductions of Shimura varieties of abelian type". arXiv: 1707.00439 (cit. on p. 1380).

Yichao Tian and Liang Xiao (Oct. 2014). "Tate cycles on some quaternionic Shimura varieties mod p". arXiv: 1410.2321 (cit. on p. 1377).

– (2016). "On Goren-Oort stratification for quaternionic Shimura varieties". *Compos. Math.* 152.10, pp. 2134–2220. MR: 3570003 (cit. on p. 1377).

Eva Viehmann (2006). "The dimension of some affine Deligne-Lusztig varieties". *Ann. Sci. École Norm. Sup. (4)* 39.3, pp. 513–526. MR: 2265677 (cit. on p. 1371).

– (2008). "Connected components of closed affine Deligne-Lusztig varieties". *Math. Ann.* 340.2, pp. 315–333. MR: 2368982 (cit. on p. 1373).
– (2014). "Truncations of level 1 of elements in the loop group of a reductive group". *Ann. of Math. (2)* 179.3, pp. 1009–1040. MR: 3171757 (cit. on pp. 1376, 1377).
– (Nov. 2015). "On the geometry of the Newton stratification". arXiv: 1511.03156 (cit. on p. 1379).
Eva Viehmann and Torsten Wedhorn (2013). "Ekedahl-Oort and Newton strata for Shimura varieties of PEL type". *Math. Ann.* 356.4, pp. 1493–1550. MR: 3072810 (cit. on p. 1379).
Inken Vollaard and Torsten Wedhorn (2011). "The supersingular locus of the Shimura variety of $\mathrm{GU}(1, n-1)$ II". *Invent. Math.* 184.3, pp. 591–627. MR: 2800696 (cit. on p. 1377).
Torsten Wedhorn (1999). "Ordinariness in good reductions of Shimura varieties of PEL-type". *Ann. Sci. École Norm. Sup. (4)* 32.5, pp. 575–618. MR: 1710754 (cit. on p. 1379).
J.-P. Wintenberger (2005). "Existence de F-cristaux avec structures supplémentaires". *Adv. Math.* 190.1, pp. 196–224. MR: 2104909 (cit. on p. 1370).
Wei Zhang (2012). "On arithmetic fundamental lemmas". *Invent. Math.* 188.1, pp. 197–252. MR: 2897697 (cit. on pp. 1364, 1377).
Rong Zhou (July 2017). "Mod-p isogeny classes on Shimura varieties with parahoric level structure". arXiv: 1707.09685 (cit. on pp. 1375, 1380).
Xinwen Zhu (2014). "On the coherence conjecture of Pappas and Rapoport". *Ann. of Math. (2)* 180.1, pp. 1–85. MR: 3194811 (cit. on p. 1366).
– (2017). "Affine Grassmannians and the geometric Satake in mixed characteristic". *Ann. of Math. (2)* 185.2, pp. 403–492. MR: 3612002 (cit. on pp. 1364, 1371).

Received 2017-12-05.

Xuhua He (何旭华)
Department of Mathematics
University of Maryland
College Park MD 20742
USA
xuhuahe@math.umd.edu

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1385–1410)

EXT-ANALOGUES OF BRANCHING LAWS

Dipendra Prasad

Abstract

We consider the Ext-analogues of branching laws for representations of a group to its subgroups in the context of p-adic groups.

Contents

1 Introduction

Considering the restriction of representations of a group G to one of its subgroups H, say of $G = \mathrm{SO}_{n+1}(F)$ to $H = \mathrm{SO}_n(F)$ for a non-archimedean local field F has been a very

MSC2010: primary 11F70; secondary 22E55.

fruitful direction of research especially through its connections to questions on period integrals of automorphic representations, cf. Gan, Gross, and Prasad [2012] for the conjectural theory both locally and globally. The question for local fields amounts to understanding $\mathrm{Hom}_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$ for irreducible admissible representations π_1 of $\mathrm{SO}_{n+1}(F)$, and π_2 of $\mathrm{SO}_n(F)$. The first result proved about this is the multiplicity one property which says that this space is at most one dimensional, cf. Aizenbud, Gourevitch, Rallis, and Schiffmann [2010], Sun and Zhu [2012]. It may be mentioned that before the full multiplicity one theorem was proved, even finite dimensionality of the space was not known. With multiplicity one theorem proved, one then goes on to prove more precise description of the set of irreducible admissible representations π_1 of $\mathrm{SO}_{n+1}(F)$, and π_2 of $\mathrm{SO}_n(F)$ with $\mathrm{Hom}_{\mathrm{SO}_n(F)}[\pi_1, \pi_2] \neq 0$. These have now become available in a series of papers due to Waldspurger, and Moeglin-Waldspurger, cf. J.-L.Waldspurger [2010], J.-L.Waldspurger [2012b], J.-L.Waldspurger [2012a], Moeglin and J.-L.Waldspurger [2012]. There is also a recent series of papers by Beuzart-Plessis on similar questions for unitary groups, cf. Beuzart-Plessis [2016], Beuzart-Plessis [2015], Beuzart-Plessis [2014].

Given the interest in the space $\mathrm{Hom}_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$, it is natural to consider the related spaces $\mathrm{Ext}^i_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$, and in fact homological algebra methods suggest that the simplest answers are not for these individual spaces, but for the alternating sum of their dimensions: $\mathrm{EP}[\pi_1, \pi_2] = \sum_{i=0}^{\infty} (-1)^i \dim \mathrm{Ext}^i_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$; these hopefully more manageable objects –certainly more flexible– when coupled with vanishing of higher Ext's (when available) may give theorems about $\mathrm{Hom}_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$. We hasten to add that before we can define $\mathrm{EP}[\pi_1, \pi_2]$, $\mathrm{Ext}^i_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$ needs to be proved to be finite dimensional for π_1 and π_2 finite length admissible representations of $\mathrm{SO}_{n+1}(F)$ and $\mathrm{SO}_n(F)$ respectively, and also proved to be 0 for i large. Vanishing of $\mathrm{Ext}^i_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$ for large i is a well-known generality to which we will come to later. Towards a proof of finite dimensionality of Ext^i in this case, to be made by an inductive argument on n later in the paper, we note that unlike $\mathrm{Hom}_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$, where we will have no idea how to prove finite dimensionality if both π_1 and π_2 are cuspidal, exactly this case we can handle apriori, for $i > 0$, as almost by the very definition of cuspidal representations, they are both projective and injective objects in the category of smooth representations. Recently, there is a very general finiteness theorem for $\mathrm{Ext}^i[\pi_1, \pi_2]$ (for spherical varieties) in Aizenbud and Sayag [2017]. However, we have preferred to give our own older approach via Bessel models which intervene when analyzing principal series representations of $\mathrm{SO}_{n+1}(F)$ when restricted to $\mathrm{SO}_n(F)$. As a bonus, this approach gives explicit answers about Euler-Poincaré characteristics.

Thinking about Ext-analogues suggest interchanging the roles of π_1 and π_2 in analogy with the known relationship, $\mathrm{EP}[V_1, V_2^\vee] = \mathrm{EP}[V_2, V_1^\vee]$ when V_1 and V_2 are finite length representations on the same group, and allows one to consider submodules as in $\mathrm{Hom}_{\mathrm{SO}_n(F)}[\pi_2, \pi_1]$, and more generally, $\mathrm{Ext}^i_{\mathrm{SO}_n(F)}[\pi_2, \pi_1]$.

Based on various examples, a clear picture seems to be emerging about $\mathrm{Ext}^i_H[\pi_1, \pi_2]$. For example, we expect that when π_1 and π_2 are tempered, $\mathrm{Ext}^i_H[\pi_1, \pi_2]$ is nonzero only for $i = 0$. On the other hand we expect that $\mathrm{Ext}^i_H[\pi_2, \pi_1]$ is typically zero for $i = 0$ (so no wonder branching is usually not considered as a subrepresentation!), and shows up only for i equals the split rank of the center of the Levi from which π_2 arises through parabolic induction of a supercuspidal representation; in fact $\mathrm{Ext}^i_H[\pi_2, \pi_1]$ is zero beyond the split rank of the center of this Levi by generalities, so $\mathrm{Ext}^i_H[\pi_1, \pi_2]$ is typically nonzero only for $i = 0$, whereas $\mathrm{Ext}^i_H[\pi_2, \pi_1]$ is nonzero only for the largest possible i. We make precise some of these suggestions during the course of the paper, and discuss some examples as evidence for the suggested conjectures made here.

In the process of relating $\mathrm{Ext}^i[\pi_1, \pi_2]$ with $\mathrm{Ext}^i[\pi_2, \pi_1]$, we were led to a duality theorem for a general reductive group which turned out to be a consequence of the work in P.Schneider and U.Stuhler [1997]. It is the subject matter of Section 8. As an example, calculation of $\dim_{\mathbb{C}} \mathrm{Hom}_{\mathrm{PGL}_2(F)}[\pi_1 \otimes \pi_2, \pi_3]$ which was part of author's work in Prasad [1990], and simple calculations about $\mathrm{EP}_{\mathrm{PGL}_2(F)}[\pi_1 \otimes \pi_2, \pi_3]$ allow the calculation of $\mathrm{Ext}^1_{\mathrm{PGL}_2(F)}[\pi_1 \otimes \pi_2, \pi_3]$, and then by the duality theorem, we are able to analyze $\mathrm{Hom}_{\mathrm{PGL}_2(F)}[\pi_3, \pi_1 \otimes \pi_2]$ (irreducible submodules of the tensor product).

In the archimedean case, several papers of T. Kobayashi, see e.g. Kobayashi [1994], do study the restriction problem for $(\mathfrak{g}, K)$-modules in the sense of sub-modules but the analogous restriction problem in the sense of sub-modules seems to be absent in the p-adic case. Proposition 5.1 and Proposition 5.3 suggest that $\mathrm{Hom}_H[\pi_2, \pi_1] = 0$ whenever π_1 is an irreducible tempered representation of G (assumed to be simple) unless H has compact center, and π_2 is a supercuspidal representation of it.

To summarize the main results of the paper, we might mention Theorem 4.2 giving a complete understanding of $\mathrm{EP}_{\mathrm{GL}_n(F)}[\pi_1, \pi_2]$ for π_1 and π_2 finite length representations of $\mathrm{GL}_{n+1}(F)$ and $\mathrm{GL}_n(F)$ respectively. Theorem 6.1 proves $\mathrm{Ext}^i_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$ to be finite dimensional for π_1 and π_2 finite length representations of $\mathrm{SO}_{n+1}(F)$ and $\mathrm{SO}_n(F)$ respectively, and as Corollary 6.3 of the proof, gives a good understanding of $\mathrm{EP}_{\mathrm{SO}_n(F)}[\pi_1, \pi_2]$ when π_1 is a principal series. We formulate as Conjecture 5.1 the vanishing of $\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\pi_1, \pi_2]$ for $i > 0$ for generic representations, and Conjecture 7.1 suggests that the integral formula discovered by Waldspurger in the papers J.-L.Waldspurger [2010] and J.-L.Waldspurger [2012b] are actually for Euler-Poincaré characteristic of general finite length representations in the spirit of Kazhdan orthogonality. In Section 9 we suggest that all nontrivial Ext's have some 'geometric' origin.

Acknowledgments. The author thanks Ann-Marie Aubert, Jeffrey Adams, U.K. Anandavardhanan, Joseph Bernstein, Atsushi Ichino, Raphael Beuzart-Plessis, Peter Schneider, Maarten Solleveld, and Sandeep Varma for useful discussions and correspondence. This paper is an updated version of the paper with the same title in arXiv from 2013!

2 Preliminaries

Given a connected reductive F-group G, we make the usual abuse of notation to also denote by G the locally compact totally disconnected group $G(F)$ of F-rational points of the algebraic group G. We denote by $\mathfrak{R}(G)$ the abelian category of smooth representations of G over $\mathbb{C}$. The abelian category $\mathfrak{R}(G)$ has enough projectives and enough injectives, e.g. for any compact open subgroup K of G, $\mathrm{ind}_K^G(\mathbb{C})$ is a projective object in $\mathfrak{R}(G)$, and $\mathrm{Ind}_K^G(\mathbb{C})$ is an injective object in $\mathfrak{R}(G)$ (we use throughout the paper ind for compactly supported induction and Ind for induction without compact support condition); in fact these projective objects and their direct summands, and their smooth duals as injective objects suffice for all considerations in the paper. Since $\mathfrak{R}(G)$ has enough projectives and enough injectives, it is meaningful to talk about $\mathrm{Ext}^i_G[\pi_1, \pi_2]$ as the derived functors of $\mathrm{Hom}_G[\pi_1, \pi_2]$.

For reductive p-adic groups G considered in this paper, it is known that $\mathrm{Ext}^i_G[\pi, \pi']$ is zero for any two smooth representations π and π' of G when i is greater than the F-split rank of G. This is a standard application of the projective resolution of the trivial representation $\mathbb{C}$ of G provided by the building associated to G. For another proof of this, and for finite dimensionality of $\mathrm{Ext}^i_G[\pi, \pi']$, see Proposition 2.9 below.

For two smooth representations π and π' of G one can consider the Euler-Poincaré pairing $\mathrm{EP}_G[\pi, \pi']$ between π and π' defined by

$$\mathrm{EP}_G[\pi, \pi'] = \sum_i (-1)^i \dim_{\mathbb{C}} \mathrm{Ext}^i_G[\pi, \pi'].$$

For this definition to make sense, we must prove that $\mathrm{Ext}^i_G[\pi, \pi']$ are finite-dimensional vector spaces over $\mathbb{C}$ for all integers i. An obvious remark which will be tacitly used throughout this paper is that if

$$0 \to \pi_1 \to \pi \to \pi_2 \to 0,$$

is an exact sequence of smooth G-modules, and if any two of the $\mathrm{EP}_G[\pi_1, \pi']$, $\mathrm{EP}_G[\pi, \pi']$, $\mathrm{EP}_G[\pi_2, \pi']$, make sense, then so does the third (finite dimensionality of the Ext groups, and zero beyond a stage), and

$$\mathrm{EP}_G[\pi, \pi'] = \mathrm{EP}_G[\pi_1, \pi'] + \mathrm{EP}_G[\pi_2, \pi'].$$

This remark will be used to break up representations π or π' in terms of simpler objects for which EP can be proved to make sense by reducing to smaller groups via some form of Frobenius recoprocity.

The following proposition summarizes some key properties of the Euler-Poincaré pairing, see P.Schneider and U.Stuhler [1997] for the proofs (part 2.1 (4) is known only in characteristic zero).

Proposition 2.1. *Let π and π' be finite-length, smooth representations of a reductive p-adic group G. Then:*

1. $\mathrm{EP}[\pi_1^\vee, \pi_2]$ *is a symmetric, $\mathbb{Z}$-bilinear form on the Grothendieck group of finite-length representations of G.*

2. EP *is locally constant. (A family $\{\pi_\lambda\}$ of representations on a fixed vector space V is said to* vary continuously *if all $\pi_\lambda|_K$ are all equivalent for some compact open subgroup K, and the matrix coefficients $\langle \pi_\lambda v, \tilde{v}\rangle$ vary continuously in λ.)*

3. $\mathrm{EP}_G[\pi, \pi'] = 0$ *if π or π' is induced from any proper parabolic subgroup in G.*

4. $\mathrm{EP}_G[\pi, \pi'] = \int_{C_{ellip}} \Theta(c)\bar{\Theta}'(c)\, dc$, *where Θ and Θ' are the characters of π and π' assumed to have the same unitary central character, and dc is a natural measure on the set C_{ellip} of regular elliptic conjugacy classes in G. (Note that if G has non-compact center, then both sides of this equality are zero; the right hand side being zero as there are no regular elliptic elements in G in that case, and the left hand side being zero by a simple argument.)*

Several assertions about Hom spaces can be converted into assertions about Ext^i. The following generality allows one to do so.

Proposition 2.2. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two abelian categories, and $\mathfrak{F}$ a functor from $\mathfrak{A}$ to $\mathfrak{B}$, and $\mathfrak{G}$ a functor from $\mathfrak{B}$ and $\mathfrak{A}$. Assume that $\mathfrak{G}$ is a left adjoint of $\mathfrak{F}$, i.e., there is a natural equivalence of functors:*

$$\mathrm{Hom}_{\mathfrak{B}}[X, \mathfrak{F}(Y)] \cong \mathrm{Hom}_{\mathfrak{A}}[\mathfrak{G}(X), Y].$$

Then,

1. *If $\mathfrak{F}$ and $\mathfrak{G}$ are exact functors, then $\mathfrak{F}$ maps injective objects of $\mathfrak{A}$ to injective objects of $\mathfrak{B}$, and $\mathfrak{G}$ maps projective objects of $\mathfrak{B}$ to projective objects of $\mathfrak{A}$.*

2. *If $\mathfrak{F}$ and $\mathfrak{G}$ are exact functors, then* $\mathrm{Ext}^i_{\mathfrak{B}}[X, \mathfrak{F}(Y)] \cong \mathrm{Ext}^i_{\mathfrak{A}}[\mathfrak{G}(X), Y]$.

Proof. Part (1) of the Proposition follows directly from definitions; see Bernstein [1992], Proposition 8. For part (2), it suffices to note that if

$$\cdots P_n \to P_{n-1} \to \cdots \to P_1 \to P_0 \to X \to 0,$$

is a projective resolution of an object X in $\mathfrak{B}$, then by part (1) of the Proposition,

$$\cdots \mathcal{G}(P_n) \to \mathcal{G}(P_{n-1}) \to \cdots \to \mathcal{G}(P_1) \to \mathcal{G}(P_0) \to \mathcal{G}(X) \to 0,$$

is a projective resolution of $\mathcal{G}(X)$. Therefore part (2) of the proposition follows from the adjointness relationship between $\mathcal{F}$ and $\mathcal{G}$. □

The following is a direct consequence of Frobenius reciprocity combined with Proposition 2.2.

Proposition 2.3. *Let H be a closed subgroup of a p-adic Lie group G. Then,*

1. *The restriction of any smooth projective representation of G to H is a projective object in $\mathfrak{R}(H)$, and $\mathrm{Ind}_H^G U$ is an injective representation of G for any injective representation U of H.*
2. *For any smooth representation U of H, and V of G,*

$$\mathrm{Ext}^i_G[V, \mathrm{Ind}_H^G U] \cong \mathrm{Ext}^i_H[V, U].$$

Note that for any two smooth representations U, V of G,

$$\mathrm{Hom}_G[U, V^\vee] \cong \mathrm{Hom}_G[V, U^\vee],$$

where $U^\vee, V^\vee$ are the smooth duals of U, V respectively. Therefore we have adjoint functors as in Proposition 2.2 with $\mathcal{F} = \mathcal{G}$ to be the smooth dual from the category of smooth representations of a p-adic group G to its opposite category. By Proposition 2.2, it follows that the smooth dual of a projective object in $\mathfrak{R}(G)$ is an injective object in $\mathfrak{R}(G)$, and further we have the following proposition.

Proposition 2.4. *For a p-adic Lie group G, let U and V be two smooth representations of G. Then,*

$$\mathrm{Ext}^i_G[U, V^\vee] \cong \mathrm{Ext}^i_G[V, U^\vee].$$

Since the smooth dual of $\mathrm{ind}_H^G(U)$ is $\mathrm{Ind}_H^G(U^\vee)$ (for normalized induction), the previous two propositions combine to give:

Proposition 2.5. *For H a closed subgroup of a p-adic Lie group G, let U be a smooth representation of H, and V a smooth representation of G. Then,*

$$\mathrm{Ext}^i_H[V, U^\vee] \cong \mathrm{Ext}^i_G[V, \mathrm{Ind}_H^G(U^\vee)] \cong \mathrm{Ext}^i_G[\mathrm{ind}_H^G U, V^\vee].$$

For smooth representations U, V, W of G, the canonical isomorphism,

$$\mathrm{Hom}_G[V \otimes U, W^\vee] \cong \mathrm{Hom}_G[U \otimes W, V^\vee],$$

translates into the following proposition by Proposition 2.2.

Proposition 2.6. *For a p-adic Lie group G, and U, V, W smooth representations of G, there are canonical isomorphisms,*

$$\mathrm{Ext}^i_G[V \otimes U, W^\vee] \cong \mathrm{Ext}^i_G[U \otimes W, V^\vee],$$

in particular

$$\mathrm{Ext}^i_G[V \otimes U, \mathbb{C}] \cong \mathrm{Ext}^i_G[V, U^\vee].$$

Proposition 2.2 with the form of Frobenius reciprocity for Jacquet modules, implies the following proposition.

Proposition 2.7. *For P a parabolic subgroup of a reductive p-adic group G with Levi decomposition $P = MN$, the Jacquet functor $V \to V_N$ from $\mathfrak{R}(G)$ to $\mathfrak{R}(M)$ takes projective objects to projective objects, and for $V \in \mathfrak{R}(G)$, $U \in \mathfrak{R}(M)$, we have (using normalized parabolic induction and normalized Jacquet module),*

$$\mathrm{Ext}^i_G[V, \mathrm{Ind}_P^G U] \cong \mathrm{Ext}^i_M[V_N, U].$$

The proof of the following proposition is exactly as the proof of the earlier proposition. This proposition will play an important role in setting-up an inductive context to prove theorems on a group G in terms of similar theorems for subgroups.

Proposition 2.8. *For P a (not necessarily parabolic) subgroup of a reductive p-adic group G with Levi decomposition $P = MN$, let ψ be a character of N normalized by M. Then for any irreducible representation μ of M, one can define a representation of P, denoted by $\mu \cdot \psi$ which when restricted to M is μ, and when restricted to N is ψ. For any smooth representation V of G, let $V_{N,\psi}$ be the twisted Jacquet module of V with respect to the character ψ of N which is a smooth representation of M. Then,*

$$\mathrm{Ext}^i_G[\mathrm{ind}_P^G(\mu \cdot \psi), V^\vee] \cong \mathrm{Ext}^i_M[V_{N,\psi}, \mu^\vee].$$

The following much deeper result than these earlier results follows from the so called *Bernstein's second adjointness theorem.*

Theorem 2.1. *For P a parabolic subgroup of a reductive p-adic group G with Levi decomposition $P = MN$, let U be a smooth representation of M thought of as a representation of P, and V a smooth representation of G. Let $P^- = MN^-$ be the parabolic* *opposite to $P = MN$. Then we have (using normalized parabolic induction and normalized Jacquet module),*

$$\mathrm{Ext}^i_G[\mathrm{Ind}_P^G U, V] \cong \mathrm{Ext}^i_M[U, V_{N^-}].$$

As a sample of arguments with the Ext groups, we give a proof of the following basic proposition.

Proposition 2.9. *Suppose that V is a smooth representation of G of finite length, and that all of its irreducible subquotients are subquotients of representations induced from supercuspidal representations of a Levi factor of the standard parabolic subgroup $P = MU$ of G, defined by a subset Θ of the set of simple roots for a maximal split torus of G. Then if V' is a finite length smooth representation of G, $\mathrm{Ext}^i_G[V, V']$ and $\mathrm{Ext}^i_G[V', V]$ are finite dimensional vector spaces over $\mathbb{C}$. If V' is any smooth representation of G, $\mathrm{Ext}^i_G[V, V'] = \mathrm{Ext}^i_G[V', V] = 0$ for $i > d(M) = d - |\Theta|$ where d is the F-split rank of G. Further, $\mathrm{EP}_G[\pi, \pi'] = 0$ if both π, π' are finite length representations of G, and π or π' is induced from a proper parabolic subgroup of G.*

Proof. This is Corollary III.3.3 in P.Schneider and U.Stuhler [1997]. Since this is elementary enough, we give another proof.

We begin by noting that tensoring V by the resolution of $\mathbb{C}$ by projective modules in $\mathfrak{R}(G)$ afforded by the building associated to G gives a projective resolution of V, but one which is not finitely generated as a G-module even if V is irreducible, and therefore proving finite dimensionality of $\mathrm{Ext}^i_G[V, V']$ requires some work. The resolution given by the building at least proves that these are 0 beyond the split rank of G. Our proof below first proves the assertions on $\mathrm{Ext}^i_G[V, V']$ if V or V' is a full principal series $\mathrm{Ind}_P^G \rho$ where ρ is a cuspidal representation of M, and then handles all subquotients by a standard *dévissage*.

Fix a surjective map $\phi : M \to \mathbb{Z}^{d(M)}$ with kernel M^ϕ which is sometimes called the subgroup of M generated by compact elements.

Let ρ be a cuspidal representation of M. Therefore ρ restricted to M^ϕ, which is $[M, M]$ up to a compact group, is an injective module, and hence $\mathrm{Ext}^i_{M^\phi}[V_N, \rho] = 0$ for $i > 0$. By Frobenius reciprocity, combined with the spectral sequence associated to the normal subgroup M^ϕ of M with quotient $\mathbb{Z}^d$, it follows that:

$$\begin{aligned}\mathrm{Ext}^i_G[V, \mathrm{Ind}_P^G \rho] &\cong \mathrm{Ext}^i_M[V_N, \rho] \\ &\cong H^i(\mathbb{Z}^{d(M)}, \mathrm{Hom}_{M^\phi}[V_N, \rho]).\end{aligned}$$

This proves that $\mathrm{Ext}^i_G[V, \mathrm{Ind}_P^G \rho] = 0$ for $i > d(M)$ for any smooth representation V of G, and that $\mathrm{Ext}^i_G[V, \mathrm{Ind}_P^G \rho]$ are finite dimensional for V of finite length.

Similarly, by the second adjointness theorem, it follows that $\mathrm{Ext}^i_G[\mathrm{Ind}_P^G(\rho), V'] = 0$ for $i > d(M)$, and that $\mathrm{Ext}^i_G[\mathrm{Ind}_P^G(\rho), V']$ are finite dimensional for V' a finite length representation in $\mathfrak{R}(G)$.

Having proved properties of $\mathrm{Ext}^i_G[V, \mathrm{Ind}_P^G \rho]$ and $\mathrm{Ext}^i_G[\mathrm{Ind}_P^G(\rho), V']$, the rest of the proposition about $\mathrm{Ext}^i_G[V, V']$ follows by *dévissage* by writing an irreducible representation V of G as a quotient of a principal series $Ps = \mathrm{Ind}_P^G \rho$, and using conclusions on the principal series to make conclusions on V. This part of the argument is very similar to what we give in Lemma 6.1, so we omit it here. □

3 Kunneth Theorem

In this section, we prove a form of the Kunneth theorem which we will have several occasions to use. A version of Kunneth's theorem is there in Raghuram [2007] assuming, however, finite length conditions on both E_1 and E_2 which is not adequate for our applications.

During the course of the proof of the Kunneth Theorem, we will need to use the following most primitive form of Frobenius reciprocity.

Lemma 3.1. *Let K be an open subgroup of a p-adic group G. Let E be a smooth representation of K, and F a smooth representation of G. Then,*

$$\mathrm{Hom}_G[\mathrm{ind}_K^G E, F] \cong \mathrm{Hom}_K[E, F].$$

Theorem 3.1. *Let G_1 and G_2 be two p-adic groups. Let E_1, F_1 be any two smooth representations of G_1, and E_2, F_2 be any two smooth representations of G_2. Then assuming that G_1 is a reductive p-adic group, and E_1 has finite length, we have*

$$\mathrm{Ext}^i_{G_1 \times G_2}[E_1 \boxtimes E_2, F_1 \boxtimes F_2] \cong \bigoplus_{i=j+k} \mathrm{Ext}^j_{G_1}[E_1, F_1] \otimes \mathrm{Ext}^k_{G_2}[E_2, F_2].$$

Proof. If P_1 is a projective module for G_1, and P_2 a projective module for G_2, then $P_1 \boxtimes P_2$ is a projective module for $G_1 \times G_2$.

Let

$$\cdots \to P_1 \to P_0 \to E_1 \to 0,$$

$$\cdots \to Q_1 \to Q_0 \to E_2 \to 0,$$

be a projective resolution for E_1 as a G_1-module, and a projective resolution for E_2 as a G_2-module.

It follows that the tensor product of these two exact sequences:

$$\cdots \to P_1 \boxtimes Q_0 + P_0 \boxtimes Q_1 \to P_0 \boxtimes Q_0 \to E_1 \boxtimes E_2 \to 0,$$

is a projective resolution of $E_1 \boxtimes E_2$. Therefore, $\mathrm{Ext}^i_{G_1\times G_2}[E_1 \boxtimes E_2, F_1 \boxtimes F_2]$ can be calculated by taking the cohomology of the chain complex $\mathrm{Hom}_{G_1\times G_2}[\bigoplus_{i+j=k} P_i \boxtimes Q_j, F_1 \boxtimes F_2]$.

It is possible to choose a projective resolution of E_1 by $P_i = \mathrm{ind}_{K_i}^{G_1} W_i$ for finite dimensional representations W_i of compact open subgroups K_i of G_1. The existence of such a projective resolution is made possible through the construction of an equivariant sheaf on the Bruhat-Tits building of G_1 associated to the representation E_1, cf. P.Schneider and U.Stuhler [1997]. this is the step which needs G_1 to be reductive, and also requires the admissibility of E_1.

Since W_i are finite dimensional, we have the isomorphism

$$\mathrm{Hom}_{K\times G_2}[W_i \boxtimes Q_j, F_1 \boxtimes F_2] \cong \mathrm{Hom}_K[W_i, F_1] \otimes \mathrm{Hom}_{G_2}[Q_j, F_2],$$

therefore,

$$\begin{aligned}\mathrm{Hom}_{G_1\times G_2}[P_i \boxtimes Q_j, F_1 \boxtimes F_2] &= \mathrm{Hom}_{G_1\times G_2}[\mathrm{ind}_K^{G_1}(W_i) \boxtimes Q_j, F_1 \boxtimes F_2] \\ &\cong \mathrm{Hom}_{K\times G_2}[W_i \boxtimes Q_j, F_1 \boxtimes F_2] \\ &\cong \mathrm{Hom}_K[W_i, F_1] \otimes \mathrm{Hom}_{G_2}[Q_j, F_2] \\ &\cong \mathrm{Hom}_{G_1}[P_i, F_1] \otimes \mathrm{Hom}_{G_2}[Q_j, F_2].\end{aligned}$$

Thus we are able to identify the chain complex $\mathrm{Hom}_{G_1\times G_2}[\bigoplus_{i+j=k} P_i \boxtimes Q_j, F_1 \boxtimes F_2]$ as the tensor product of the chain complexes $\mathrm{Hom}_{G_1}[P_i, F_1]$ and $\mathrm{Hom}_{G_2}[Q_j, F_2]$. Now the abstract Kunneth theorem which calculates the cohomology of the tensor product of two chain complexes in terms of the cohomology of the individual chain complexes completes the proof of the theorem. □

4 Branching laws from $\mathrm{GL}_{n+1}(F)$ to $\mathrm{GL}_n(F)$

We begin by recalling the following basic result in this context, cf. Prasad [1993].

Theorem 4.1. *Given an irreducible generic representation π_1 of $\mathrm{GL}_{n+1}(F)$, and an irreducible generic representation π_2 of $\mathrm{GL}_n(F)$,*

$$\mathrm{Hom}_{\mathrm{GL}_n(F)}[\pi_1, \pi_2] = \mathbb{C}.$$

The aim of this section is to prove the following theorem which can be considered as the Euler-Poincaré version of Theorem 4.1.

Theorem 4.2. *Let π_1 be an admissible representation of $\mathrm{GL}_{n+1}(F)$ of finite length, and π_2 an admissible representation of $\mathrm{GL}_n(F)$ of finite length. Then, $\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\pi_1, \pi_2]$ are* *finite dimensional vector spaces over* $\mathbb{C}$, *and*

$$\mathrm{EP}_{\mathrm{GL}_n(F)}[\pi_1,\pi_2] = \dim \mathrm{Wh}(\pi_1)\cdot \dim \mathrm{Wh}(\pi_2),$$

where $\mathrm{Wh}(\pi_1)$, *resp.* $\mathrm{Wh}(\pi_2)$, *denotes the space of Whittaker models for* π_1, *resp.* π_2, *with respect to fixed non-degenerate characters on the maximal unipotent subgroups in* $\mathrm{GL}_{n+1}(F)$ *and* $\mathrm{GL}_n(F)$.

The proof of this theorem will be accomplished using some results of Bernstein and Zelevinsky regarding the structure of representations of $\mathrm{GL}_{n+1}(F)$ restricted to the mirabolic subgroup.

Denote by E_n the mirabolic subgroup of $\mathrm{GL}_{n+1}(F)$ consisting of matrices whose last row is equal to $(0,0,\cdots,0,1)$ and let N_{n+1} be the group of upper triangular unipotent matrices in $\mathrm{GL}_{n+1}(F)$. We will be using subgroups $\mathrm{GL}_i(F)$ of $\mathrm{GL}_{n+1}(F)$ for $i\le n+1$ always sitting at the upper left corner of $\mathrm{GL}_{n+1}(F)$. We fix a nontrivial character ψ_0 of F and let ψ_{n+1} be the character of N_{n+1} given by

$$\psi_{n+1}(u) = \psi_0(u_{1,2}+u_{2,3}+\cdots+u_{n,n+1}).$$

For a representation π of $\mathrm{GL}_{n+1}(F)$, let

$$\pi^i = \text{ the } i\text{-th derivative of } \pi,$$

which is a representation of $\mathrm{GL}_{n+1-i}(F)$. It will be important for us to note that π^i are representations of finite length of $\mathrm{GL}_{n+1-i}(F)$ if π is of finite length for $\mathrm{GL}_{n+1}(F)$.

To recall the definition of π^i, let $R_{n+1-i} = \mathrm{GL}_{n+1-i}(F)\cdot V_i$ be the subgroup of $\mathrm{GL}_{n+1}(F)$ consisting of matrices

$$\begin{pmatrix} g & v \\ 0 & z \end{pmatrix}$$

with $g\in \mathrm{GL}_{n+1-i}(F)$, $v\in M_{(n+1-i)\times i}$, $z\in N_i$. If the character ψ_i of N_i is extended to V_i by extending it trivially across $M_{(n+1-i)\times i}$, then we have

$$\pi^i = \pi_{V_i,\psi_i},$$

where π_{V_i,ψ_i} is the twisted Jacquet module of π, i.e., the maximal quotient of π on which V_i operates via the character ψ_i.

Here is a generality from Bernstein and Zelevinsky [1976], §3.5.

Proposition 4.1. *Any smooth representation* Σ *of* E_n *has a natural filtration by* $E = E_n$ *modules*

$$0=\Sigma_0\subset\Sigma_1\subset\Sigma_2\subset\cdots\subset\Sigma_{n+1}=\Sigma$$

such that

$$\Sigma_{i+1}/\Sigma_i = \mathrm{ind}_{R_i}^{E_n}(\Sigma^{n+1-i} \otimes \psi_{n+1-i}) \quad \textit{for } i = 0, \cdots, n,$$

where $R_i = \mathrm{GL}_i(F) \cdot V_{n+1-i}$ *is the subgroup of* $\mathrm{GL}_{n+1}(F)$ *consisting of*

$$\begin{pmatrix} g & v \\ 0 & z \end{pmatrix}$$

with $g \in \mathrm{GL}_i(F)$, $v \in M_{i\times(n+1-i)}$, $z \in N_{n+1-i}$, *and the character* ψ_{n+1-i} *on* N_{n+1-i} *is extended to* V_{n+1-i} *by extending it trivially across* $M_{i\times(n+1-i)}$.

The proof of the following proposition is a direct consequence of Proposition 2.8.

Proposition 4.2. *For a smooth representation* π_1 *of* $\mathrm{GL}_{n+1}(F)$, *and* π_2 *of* $\mathrm{GL}_n(F)$,

$$\mathrm{Ext}^j_{\mathrm{GL}_n(F)}[\mathrm{ind}_{R_i}^{\mathrm{GL}_n(F)}(\pi_1^{n-i+1} \otimes \psi_{n-i}), \pi_2^\vee] = \mathrm{Ext}^j_{\mathrm{GL}_i(F)}[\pi_2^{n-i}, (\pi_1^{n-i+1})^\vee].$$

Proof. Since $\mathrm{GL}_n(F) \cdot R_i = E_n$, $\mathrm{GL}_n(F) \cap R_i = \mathrm{GL}_i(F) \cdot V_{n-i}$, the restriction of $\pi_1^{n+1-i} \otimes \psi_{n+1-i}$ from R_i to $\mathrm{GL}_n(F) \cap R_i = \mathrm{GL}_i(F) \cdot V_{n-i}$ is $\pi_1^{n+1-i} \otimes \psi_{n-i}$ for any i, $0 \le i \le n$. Therefore, the proposition follows from Proposition 2.8. □

Lemma 4.1. *For any two smooth representations* V_1, V_2 *of* $\mathrm{GL}_n(F)$, $n \ge 1$, *of finite length,*

$$\mathrm{EP}_{\mathrm{GL}_n(F)}[V_1, V_2] = 0.$$

Proof. It suffices to prove the lemma assuming that both V_1 and V_2 are irreducible representations of $\mathrm{GL}_n(F)$. If the two representations V_1, V_2 were irreducible, and had different central characters, then clearly $\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[V_1, V_2] = 0$ for all integers i. On the other hand, we know by Proposition 2.1(b) that for representations V_1 and V_2 of finite length of $\mathrm{GL}_n(F)$, $\mathrm{EP}_{\mathrm{GL}_n(F)}[V_1, V_2]$ is constant in a connected family, so denoting by ν the character $\nu(g) = |\det g|$ of $\mathrm{GL}_n(F)$, we have $\mathrm{EP}_{\mathrm{GL}_n(F)}[V_1, V_2] = \mathrm{EP}_{\mathrm{GL}_n(F)}[\nu^s \cdot V_1, V_2]$ for all $s \in \mathbb{C}$. Choosing s appropriately, we can change the central character of $\nu^s \cdot V_2$ to be different from V_1, and hence $\mathrm{EP}_{\mathrm{GL}_n(F)}[V_1, V_2] = \mathrm{EP}_{\mathrm{GL}_n(F)}[\nu^s \cdot V_1, V_2] = 0$. □

Proof of Theorem 4.2: Since the Euler-Poincaré characteristic is additive in exact sequences, it suffices to calculate $\mathrm{EP}_{\mathrm{GL}_n(F)}[\mathrm{ind}_{R_i}^{\mathrm{GL}_n(F)}(\pi_1^{n-i} \otimes \psi_{n-i+1}), \pi_2]$ which by Proposition 2.8 is $\mathrm{EP}_{\mathrm{GL}_i(F)}[(\pi_2^{n-i})^\vee, (\pi_1^{n-i+1})^\vee]$, which by Lemma 4.1 above is 0 unless $i = 0$. (Note that in $\mathrm{EP}_{\mathrm{GL}_i(F)}[(\pi_2^{n-i})^\vee, (\pi_1^{n-i+1})^\vee]$, both the representations involved are admissible representations of $\mathrm{GL}_i(F)$.) For $i = 0$, note that we are dealing with $\mathrm{GL}_0(F) = 1$, and the representations involved are $(\pi_1^\vee)^{n+1}$ and $(\pi_2^\vee)^n$, which are nothing but the space of Whittaker models of $\pi_1^\vee$ and $\pi_2^\vee$. Since for representations V_1, V_2 of the group $\mathrm{GL}_0(F) = 1$, $\mathrm{EP}[V_1, V_2] = \dim \mathrm{Hom}[V_1, V_2] = \dim V_1 \cdot \dim V_2$, this completes the proof of the theorem.

Remark 4.1. One knows, cf. Prasad [1993], that there are irreducible generic representations of $\mathrm{GL}_3(F)$ which have the trivial representation of $\mathrm{GL}_2(F)$ as a quotient; similarly, there are irreducible nongeneric representations of $\mathrm{GL}_3(F)$ with irreducible generic representations of $\mathrm{GL}_2(F)$ as a quotient. For such pairs (π_1, π_2) of representations, it follows from Theorem 4.2 that $\mathrm{EP}_{\mathrm{GL}_2(F)}[\pi_1, \pi_2] = 0$, whereas $\mathrm{Hom}_{\mathrm{GL}_2(F)}[\pi_1, \pi_2] \neq 0$. Therefore, for such pairs (π_1, π_2) of irreducible representations, we must have $\mathrm{Ext}^i_{\mathrm{GL}_2(F)}[\pi_1, \pi_2] \neq 0$, for some $i > 0$.

5 Conjectural vanishing of Ext groups for generic representations

The following conjecture seems to be at the root of why the simple and general result of previous section on Euler-Poincaré characteristic translates into a simple result about Hom spaces for generic representations. The author has not managed to prove it in any generality. There is a recent preprint by Chan and Savin, cf. Chan and Savin [2017] dealing with some cases of this conjecture.

Conjecture 5.1. *Let π_1 be an irreducible generic representation of $\mathrm{GL}_{n+1}(F)$, and π_2 an irreducible generic representation of $\mathrm{GL}_n(F)$. Then,*

$$\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\pi_1, \pi_2] = 0, \quad \text{for all} \quad i > 0.$$

Remark 5.1. By Remark 4.1, one cannot remove the genericity condition for either π_1 or π_2 in the above conjecture. In particular, one cannot expect that a generic representation of $\mathrm{GL}_{n+1}(F)$ when restricted to $\mathrm{GL}_n(F)$ is a projective representation in $\mathcal{R}(\mathrm{GL}_n(F))$ although this is the case for supercuspidal representations of $\mathrm{GL}_{n+1}(F)$. The paper Chan and Savin [ibid.] proves that the part of the Steinberg representation of $\mathrm{GL}_{n+1}(F)$ (denoted St_{n+1}) in the Iwahori component of the Bernstein decomposition for $\mathcal{R}(GL_n(F))$ is a projective module. There is no doubt then that St_{n+1} when restricted to $\mathrm{GL}_n(F)$ is a projective representation in $\mathcal{R}(\mathrm{GL}_n(F))$, therefore

$$\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\mathrm{St}_{n+1}, \pi_2] = 0 \quad \text{for} \quad i > 0$$

for any irreducible representation π_2 of $\mathrm{GL}_n(F)$. As a consequence, it will follow from the duality theorem of Schneider-Stuhler, cf. Theorem 8.1 below, that St_{n+1} contains no irreducible submodule of $\mathrm{GL}_n(F)$.

Towards checking the validity of this conjecture in some cases, note that by Theorem 4.1 and Theorem 4.2, under the hypothesis of the conjecture,

$$\dim \mathrm{Hom}_{\mathrm{GL}_n(F)}[\pi_1, \pi_2] = 1, \quad \text{and} \quad \mathrm{EP}[\pi_1, \pi_2] = 1.$$

It follows that if we already knew that $\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\pi_1,\pi_2]=0,\ \ i>1$, then we will also know that, $\mathrm{Ext}^1_{\mathrm{GL}_n(F)}[\pi_1,\pi_2]=0$, and the conjecture will be proved for such representations.

It is easy to see that if π_1 or π_2 is cuspidal, then $\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\pi_1,\pi_2]=0$ for $i>1$. We do one slightly less obvious case when π_1 arises as a subquotient of a principal series representation induced from a cuspidal representation of a maximal parabolic in $\mathrm{GL}_{n+1}(F)$.

It follows from Corollary III.3.3(i) of P.Schneider and U.Stuhler [1997] that if π_1 arises from a cuspidal representation of a maximal parabolic in $\mathrm{GL}_{n+1}(F)$, it has a projective resolution of length 1 in the category $\mathcal{R}_\chi(G)$ of smooth representations of $G=\mathrm{GL}_{n+1}(F)$ with central character χ. It is easy to see that a projective module in $\mathcal{R}_\chi(\mathrm{GL}_{n+1}(F))$ when considered as a representation of $\mathrm{GL}_n(F)$ is a projective module in $\mathcal{R}(\mathrm{GL}_n(F))$, cf. Proposition 3.2 in M.Nori and Prasad [2017]. This proves vanishing of $\mathrm{Ext}^i_{\mathrm{GL}_n(F)}[\pi_1,\pi_2]=0$ for $i>1$, hence also of $\mathrm{Ext}^1_{\mathrm{GL}_n(F)}[\pi_1,\pi_2]$.

This takes care of $G=\mathrm{GL}_{n+1}(F)$ for $n+1\le 3$, except that for $\mathrm{GL}_3(F)$ if both π_1 and π_2 arise as components of principal series representations induced from their Borel subgroups then there is a possibility of having nontrivial $\mathrm{Ext}^2_{\mathrm{GL}_2(F)}[\pi_1,\pi_2]$. By the duality theorem of Schneider-Stuhler, cf. Theorem 8.1 below, we will have, $\mathrm{Hom}_{\mathrm{GL}_2(F)}[D\pi_2,\pi_1]\neq 0$ (where $D\pi_2$ is the Aubert-Zelevinsky involution of π_2). The following proposition takes care of this.

Proposition 5.1. *Let π_1 be an irreducible generic representation of $\mathrm{GL}_3(F)$, and π_2 any irreducible representation of $\mathrm{GL}_2(F)$ which is not a twist of the Steinberg representation of $\mathrm{GL}_2(F)$. Then*

$$\mathrm{Hom}_{\mathrm{GL}_2(F)}[\pi_2,\pi_1]=0.$$

We will not prove this proposition here but discuss two propositions which deal with all but a few cases of the proposition above. The cases left out by the next two propositions can be handled by the Mackey restriction of an explicit principal series (especially using that different inducing data can give rise to the same principal series).

Proposition 5.2. *Let $H_1\subset H$ be p-adic groups with $Z=F^\times$ contained in the center of H with $Z\cap H_1=\{1\}$. Suppose μ is a smooth representation of H_1, and π_2 an irreducible admissible representation of H. Then $\mathrm{Hom}_H[\pi_2,\mathrm{ind}_{H_1}^H(\mu)]=0$.*

Proof. Note that for each $x\in H/H_1$, restriction of functions from H to $(Z=F^\times)\cdot x$ gives rise to $F^\times$-equivariant maps

$$\mathrm{ind}_{H_1}^H(\mu)\longrightarrow \mathcal{S}(F^\times),$$

which can be assumed to be nonzero for any $f\in\mathrm{ind}_{H_1}^H(\mu)$ by choosing $x\in H/H_1$ appropriately.

Therefore if $\mathrm{Hom}_H[\pi_2, \mathrm{ind}_{H_1}^H(\mu)] \neq 0$ choosing $x \in H/H_1$ appropriately, we get a nonzero map from $\pi_2 \longrightarrow \mathcal{S}(F^\times)$ which is ω-equivariant where ω is the central character of π_2. Since $\mathcal{S}(F^\times)$ has no functions on which $F^\times$ operates by a character, the proof of the proposition is complete. □

This lemma when combined with the Bernstein-Zelevinsky filtration in Proposition 4.1 has the following as an immediate consequence.

Proposition 5.3. *Let π_1 be any smooth representation of $\mathrm{GL}_{n+1}(F)$ of finite length, and π_2 any irreducible representation of $\mathrm{GL}_n(F)$. Then if*

$$\mathrm{Hom}_{\mathrm{GL}_n(F)}[\pi_2, \pi_1] \neq 0,$$

then π_2 appears a submodule of $J_{n,1}(\pi_1)$ where $J_{n,1}$ denotes the un-normalized Jacquet module with respect to the $(n,1)$ parabolic in $\mathrm{GL}_{n+1}(F)$ considered as a module for $\mathrm{GL}_n(F) \subset \mathrm{GL}_n(F) \times \mathrm{GL}_1(F)$; in particular, if $J_{n,1}(\pi_1) = 0$, then there are no nonzero $\mathrm{GL}_n(F)$-submodules in π_1.

Proof. The proof of the proposition is an immediate consequence of the observation that the Bernstein-Zelevinsky filtration in Proposition 4.1 when restricted to $\mathrm{GL}_n(F)$ gives rise to representations of $\mathrm{GL}_n(F)$ induced from subgroups $H_i \subset \mathrm{GL}_n(F)$ with $H_i \cap \{Z(\mathrm{GL}_n(F)) = F^\times\} = \{1\}$ except in the case when $H_i = \mathrm{GL}_n(F)$ which corresponds to the Jacquet module $J_{n,1}(\pi_1)$. □

6 Finite dimensionality of Ext groups

In this section we prove the finite dimensionality of Ext-groups in the case of $\mathrm{SO}_n(F) \subset \mathrm{SO}_{n+1}(F)$. The proof will have an inductive structure, and will involve Bessel models in the inductive step, so we begin by recalling the concept of Bessel models.

Let $V = X + D + W + Y$ be a quadratic space over the non-archimedean local field F with X and Y totally isotropic subspaces of V in duality with each other under the underlying bilinear form, D an anisotropic line in V, and W a quadratic subspace of V. Suppose that the dimension of X is k; fix a complete flag $\langle e_1 \rangle \subset \langle e_1, e_2 \rangle \subset \cdots \subset \langle e_1, e_2, \cdots, e_k \rangle = X$ of isotropic subspaces in X. Let $P = MU$ be the parabolic subgroup in $\mathrm{SO}(V)$ stabilizing this flag, with $M = \mathrm{GL}_1(F)^k \times \mathrm{SO}(D+W)$. For $W \subset V$ a codimension $2k+1$ subspace as above, the subgroup $\mathrm{SO}(W) \cdot U$ which is uniquely defined up to conjugacy by $\mathrm{SO}(V)$ makes frequent appearance in this work, as well as in other works on classical groups. We call this subgroup as the *Bessel subgroup*, and denote it as $\mathrm{Bes}(V, W) = \mathrm{SO}(W) \cdot U$.

Let $P_X = M_X \cdot U_X$ be the maximal parabolic of $\mathrm{SO}(V)$ stabilizing X. We have $M_X \cong \mathrm{GL}(X) \cdot \mathrm{SO}(W + D)$, and U_X sits in the exact sequence,

$$1 \to \Lambda^2 X \to U_X \to X \otimes (D + W) \to 1.$$

Let $\ell : U \to F$ be a linear form such that

1. its restriction to each of the simple root spaces in $\mathrm{GL}(X)$ defined by the flag $\langle e_1 \rangle \subset \langle e_1, e_2 \rangle \subset \cdots \subset \langle e_1, e_2, \cdots, e_k \rangle = X$ of isotropic subspaces in X is non-trivial;
2. its restriction to the unipotent radical of the parabolic $P_X = M_X U_X$ in $\mathrm{SO}(V)$ stabilizing X is trivial on the subgroup of U_X which is $\Lambda^2 X$;
3. and on the quotient of U_X by $\Lambda^2 X$ which can be identified to $(D + W) \otimes X$, ℓ is given by the tensor product of a linear form on $D + W$ which is trivial on W, and a linear form on X which is trivial on the subspace $\langle e_1, e_2, \cdots, e_{k-1} \rangle$.

Composing the linear form $\ell : U \to F$ with a nontrivial character $\psi_0 : F \to \mathbb{C}^\times$, we get a character $\psi : U \to \mathbb{C}^\times$. This character $\psi : U \to \mathbb{C}^\times$ depends only on $W \subset V$ a nondegenerate subspace of V of odd codimension, such that the quadratic space V/W is split, and is independent of all choices made along the way (including that of the character ψ_0). The character ψ of U is invariant under $\mathrm{SO}(W)$. For any representation σ of $\mathrm{SO}(W)$, $\mathrm{Bes}(V, W) = \mathrm{SO}(W) \cdot U$ comes equipped with the representation which is σ on $\mathrm{SO}(W)$, and ψ on U; since ψ is fixed when considering representations of $\mathrm{Bes}(V, W) = \mathrm{SO}(W) \cdot U$, we denote this representation of $\mathrm{Bes}(V, W) = \mathrm{SO}(W) \cdot U$ as σ itself or sometimes as $\sigma \otimes \psi$.

The Bessel models of a smooth representation π of $\mathrm{SO}(V)$ are irreducible admissible representations σ of $\mathrm{SO}(W)$ such that the following isomorphic spaces are nonzero:

$$\begin{aligned}\mathrm{Hom}_{\mathrm{Bes}(V,W)}[\pi, \sigma] &\cong \mathrm{Hom}_{\mathrm{SO}(V)}\left[\pi, \mathrm{Ind}_{\mathrm{Bes}(V,W)}^{\mathrm{SO}(V)}(\sigma)\right] \\ &\cong \mathrm{Hom}_{\mathrm{SO}(V)}\left[\mathrm{ind}_{\mathrm{Bes}(V,W)}^{\mathrm{SO}(V)}(\sigma^\vee), \pi^\vee\right].\end{aligned}$$

When W is a codimension one subspace of V, then $\mathrm{Bes}(V, W) = \mathrm{SO}(W)$, and the notion of a Bessel model is simply that of restriction from $\mathrm{SO}(V)$ to $\mathrm{SO}(W)$, whereas when $\dim(W) = 0, 1$, then the notion of a Bessel model is nothing but that of the Whittaker model (for a particular character of the maximal unipotent subgroup of $\mathrm{SO}(V)$ if $\dim(W) = 1$).

We can define the higher Ext versions of the Bessel models as one of the following isomorphic spaces:

$$\mathrm{Ext}^i_{\mathrm{Bes}(V,W)}[\pi,\sigma] \cong \mathrm{Ext}^i_{\mathrm{SO}(V)}\left[\pi, \mathrm{Ind}^{\mathrm{SO}(V)}_{\mathrm{Bes}(V,W)}(\sigma)\right]$$
$$\cong \mathrm{Ext}^i_{\mathrm{SO}(V)}\left[\mathrm{ind}^{\mathrm{SO}(V)}_{\mathrm{Bes}(V,W)}(\sigma^\vee), \pi^\vee\right].$$

The following proposition whose proof we will omit is analogous to that of Theorem 15.1 of Gan, Gross, and Prasad [2012]. It allows one to prove finite dimensionality of $\mathrm{Ext}^i_{\mathrm{Bes}(V,W_0)}[\pi,\sigma]$ if we know the finite dimensionality of $\mathrm{Ext}^i_{\mathrm{SO}(W)}[\pi,\sigma']$ where W is a codimension one subspace in the quadratic space V, and σ' an irreducible representation of $\mathrm{SO}(W)$.

Proposition 6.1. *Let $W \subset V$ be a nondegenerate quadratic subspace of codimension 1 over a non-archimedean local field F. Suppose that*

$$W = Y_k \oplus W_0 \oplus Y_k^\vee$$

and

$$V = Y_k \oplus V_0 \oplus Y_k^\vee,$$

with Y_k and $Y_k^\vee$ isotropic subspaces and $W_0 \subset V_0$ nondegenerate quadratic spaces with W_0 a subspace of codimension one in V_0. Let $P_W(Y_k)$ be the parabolic in $\mathrm{SO}(W)$ stabilizing Y_k with Levi subgroup

$$M = \mathrm{GL}(Y_k) \times \mathrm{SO}(W_0)$$

For an irreducible supercuspidal representation τ of $\mathrm{GL}(Y_k)$ and an irreducible admissible representation π_0 of $\mathrm{SO}(W_0)$, let

$$\tau \rtimes \pi_0 = \mathrm{Ind}^{\mathrm{SO}(W)}_{P_W(Y_k)}(\tau \boxtimes \pi_0)$$

be the corresponding (un-normalized) principal series representation of $\mathrm{SO}(W)$. Let π be an irreducible admissible representation $\mathrm{SO}(V)$ which does not belong to the Bernstein component associated to $(\mathrm{GL}(Y_k) \times \mathrm{SO}(V_0), \tau \boxtimes \mu)$ for any irreducible representation μ of $\mathrm{SO}(V_0)$. Then

$$\mathrm{Ext}^i_{\mathrm{SO}(W)}[\pi, \tau \rtimes \pi_0] \cong \mathrm{Ext}^i_{\mathrm{Bes}(V,W_0)}[\pi,\pi_0].$$

Corollary 6.1. *With the notation as above, if $\mathrm{Ext}^i_{\mathrm{SO}(W)}[\pi, \tau \rtimes \pi_0]$ are finite dimensional, then so are $\mathrm{Ext}^i_{\mathrm{Bes}(V,W_0)}[\pi,\pi_0]$.*

Proof. It suffices to observe that given π_0, there is a representation π of $\mathrm{SO}(V)$ which does not belong to the Bernstein component associated to $(\mathrm{GL}(Y_k) \times \mathrm{SO}(V_0), \tau \boxtimes \mu)$ for any irreducible representation μ of $\mathrm{SO}(V_0)$. □

We now come to the proof of finite dimensionality of the Ext groups.

Theorem 6.1. *Let $V = X + D + W + Y$ be as at the beginning of the section, a quadratic space over the non-archimedean local field F with W a quadratic subspace of codimension $2k+1$. Then for any irreducible admissible representation π of $\mathrm{SO}(V)$ and irreducible admissible representation σ of $\mathrm{SO}(W)$, $\mathrm{Ext}^i_{\mathrm{Bes}(V,W)}[\pi,\sigma]$ are finite dimensional vector spaces over $\mathbb{C}$ for all $i \geq 0$.*

Proof. The proof of this theorem will be by induction on the dimension of V. We thus assume that for any quadratic spaces $\mathcal{W} \subset \mathcal{V}$ with $\dim(\mathcal{V}) < \dim(V)$ (with $\mathcal{V}/\mathcal{W}$ a split quadratic space of odd dimension), and for any irreducible admissible representation π of $\mathrm{SO}(\mathcal{V})$ and irreducible admissible representation σ of $\mathrm{SO}(\mathcal{W})$,

$$\mathrm{Ext}^i_{\mathrm{Bes}(\mathcal{V},\mathcal{W})}[\pi,\sigma]$$

are finite dimensional vector spaces over $\mathbb{C}$ for all $i \geq 0$.

We begin by proving the theorem for a principal series representation of $\mathrm{SO}(V)$ induced from an irreducible representation of a maximal parabolic subgroup. By the previous proposition, we need only prove the finite dimensionality of $\mathrm{Ext}^i_{\mathrm{SO}(V')}[\pi,\pi']$ where V' is a codimension one subspace of V, and π' is an irreducible, admissible representation of $\mathrm{SO}(V')$.

Much of the proof below closely follows the paper Moeglin and J.-L.Waldspurger [2012], where they have to do much harder work to precisely analyze $\mathrm{Hom}_{\mathrm{SO}(V')}[\pi,\pi']$.

Assume that the dimension of V is $n+1$, and that V' is a subspace of dimension n. Let $V = X + V_0 + Y$ with X and Y totally isotropic subspaces of V of dimension m, and in perfect pairing with each other. Let P be the maximal parabolic subgroup of $\mathrm{SO}(V)$ stabilizing X. Let $M = \mathrm{GL}(X) \times \mathrm{SO}(V_0)$ be a Levi subgroup of P, $\pi_0 \otimes \sigma_0$ an irreducible representation of M realized on the space $E_{\pi_0} \otimes E_{\sigma_0}$, and $\pi = \pi_0 \rtimes \sigma_0$ the corresponding principal series representation of $\mathrm{SO}(V)$. Denote by E_π the space of function on $\mathrm{SO}(V)$ with values in $E_{\pi_0} \otimes E_{\sigma_0}$ verifying the usual conditions under left translation by $P(F)$ for defining the principal series representation $\pi = \pi_0 \rtimes \sigma_0$ of $\mathrm{SO}(V)$.

To understand the restriction of the principal series $\pi = \pi_0 \rtimes \sigma_0$ to $\mathrm{SO}(V')$, we need to analyze the orbits of $\mathrm{SO}(V')$ on $P(F)\backslash \mathrm{SO}(V)$. To every $g \in P(F)\backslash \mathrm{SO}(V)$, one can associate an isotropic subspace $g^{-1}(X)$ of V. Let $\mathcal{U}$ be the set of $g \in P(F)\backslash \mathrm{SO}(V)$ such that $\dim(g^{-1}(X) \cap V') = m-1$, and let $\mathcal{X}$ be the set of $g \in P(F)\backslash \mathrm{SO}(V)$ such that $\dim(g^{-1}(X) \cap V') = m$. Then $\mathcal{U}$ is an open subset of $P(F)\backslash \mathrm{SO}(V)$ which is a single orbit under $\mathrm{SO}(V')$, and $\mathcal{X}$ is a closed subset of $P(F)\backslash \mathrm{SO}(V)$ which is a single orbit under $\mathrm{SO}(V')$ unless n is even, and $n = 2m$ in which case there are two orbits in $\mathcal{X}$ under $\mathrm{SO}(V')$.

Denote by $E_{\pi,\mathcal{U}}$ the subspace of functions in E_π with support in $\mathcal{U}$, and denote by $E_{\pi,\mathcal{X}}$ the space $E_\pi/E_{\pi,\mathcal{U}}$. The spaces $E_{\pi,\mathcal{U}}$ and $E_{\pi,\mathcal{X}}$ are invariant under $\mathrm{SO}(V')$, and we have an exact sequence of $\mathrm{SO}(V')$-modules,

$$0 \to E_{\pi,\mathfrak{U}} \to E_\pi \to E_{\pi,\mathfrak{X}} \to 0.$$

To prove the finite dimensionality of Ext groups $\mathrm{Ext}^i_{\mathrm{SO}(V')}[E_\pi, \pi']$, it suffices to prove similar finite dimensionality theorems for the Ext groups involving the $\mathrm{SO}(V')$-modules $E_{\pi,\mathfrak{U}}$ and $E_{\pi,\mathfrak{X}}$. We analyze the two terms separately.

For analyzing $E_{\pi,\mathfrak{X}}$, we assume (after conjugation by $\mathrm{SO}(V)$) that both X and Y are contained in V'. Thus, $V' = X + V_0' + Y$ with $V_0' = V' \cap V_0$. It can be seen that,

$$E_{\pi,\mathfrak{X}} = \pi_0|.|^{1/2} \times \sigma_0|_{\mathrm{SO}(V_0')}.$$

By Theorem 2.1 (the second adjointness theorem of Bernstein),

$$\mathrm{Ext}^i_{\mathrm{SO}(V')}[E_{\pi,\mathfrak{X}}, \pi'] = \mathrm{Ext}^i_M[\pi_0|.|^{1/2} \boxtimes \sigma_0|_{\mathrm{SO}(V_0')}, \pi'_{N^-}],$$

where $M = \mathrm{GL}(X) \times \mathrm{SO}(V_0')$.

The proof of the finite dimensionality of $\mathrm{Ext}^i_{\mathrm{SO}(V')}[E_{\pi,\mathfrak{X}}, \pi']$ now follows from the induction hypothesis according to which the theorem was supposed to be known for $\mathrm{SO}(V_0') \subset \mathrm{SO}(V_0)$, besides the fact that π'_{N^-}, the Jacquet module with respect to the opposite parabolic $P^- = MN^-$ is an admissible representation of $M = \mathrm{GL}(X) \times \mathrm{SO}(V_0')$, hence has a finite filtration by tensor product of irreducible representations of $\mathrm{GL}(X)$ and $\mathrm{SO}(V_0')$, and then an application of the Kunneth theorem.

We now move on to $E_{\sigma,\mathfrak{U}}$. In this case, after conjugation by $\mathrm{SO}(V)$, we will be in the situation,

$$V' = X' + D' + V_0 + Y',$$

where X' and Y' are totally isotropic subspaces of V' of dimension $(m-1)$, and D', V_0, $X' + Y'$ are non-degenerate quadratic spaces. Let $X' = \{e_1, \cdots, e_{m-1}\}$, and $Y' = \{f_1, \cdots, f_{m-1}\}$. Let $V_k' = X_{k-1}' + D' + V_0 + Y_{k-1}'$, where $X_{k-1}' = \{e_{m-k+1}, \cdots, e_{m-1}\}$, and $Y_{k-1}' = \{f_{m-k+1}, \cdots, f_{m-1}\}$, each of dimension $(k-1)$ for $k = 1, \cdots, m$, and let $G_k' = \mathrm{SO}(V_k')$.

Using the filtration of the representation π_0 of $\mathrm{GL}(X)$ restricted to its mirabolic subgroup given by Proposition 4.1 in terms of the derivatives $\pi_0^k = \Delta^k(\pi_0)$, Moeglin and J.-L.Waldspurger [ibid.] obtain a filtration,

$$0 = \mu_{m+1} \subset \mu_m \subset \mu_{m-1} \cdots \subset \mu_1 = E_{\sigma,\mathfrak{U}},$$

with,

$$\mu_k/\mu_{k+1} \cong \Delta^k(\pi_0) \rtimes \mu_k',$$

as modules for $\mathrm{SO}(V')$; the representation $\Delta^k(\pi_0) \rtimes \mu'_k$ is a principal series representation of $\mathrm{SO}(V')$ induced from a parabolic $P_k = M_k N_k$ with Levi subgroup $M_k = GL_{m-k}(F) \times G'_k$, and where

$$\mu'_k = \mathrm{ind}_{\mathrm{Bes}(V'_k, V_0)}^{G'_k}(\sigma_0),$$

By Theorem 2.1,

$$\mathrm{Ext}^i_{\mathrm{SO}(V')}[\mu_k/\mu_{k+1}, \pi'] = \mathrm{Ext}^i_{\mathrm{SO}(V')}[\Delta^k(\pi_0) \rtimes \mu'_k, \pi'] \cong \mathrm{Ext}^i_{M_k}[\Delta^k(\pi_0) \boxtimes \mu'_k, \pi'_{N_k^-}],$$

with $M_k = GL_{m-k}(F) \times G'_k$, a Levi subgroup in $\mathrm{SO}(V')$. Once again Kunneth theorem implies the finite dimensionality of the Ext groups,

$$\mathrm{Ext}^i_{M_k}[\Delta^k(\pi_0) \boxtimes \mu'_k, \pi'_{N_k^-}].$$

Having proved the theorem for principal series representations of $\mathrm{SO}(V)$ induced from maximal parabolics, the next lemma proves the theorem in general. □

Lemma 6.1. *Let V be a quadratic space over the non-archimedean local field F with W a quadratic subspace of codimension 1. If for any principal series representation Ps of $\mathrm{SO}(V)$ induced from a maximal parabolic and any irreducible admissible representation σ of $\mathrm{SO}(W)$, $\mathrm{Ext}^i_{\mathrm{SO}(W)}[Ps, \sigma]$ are finite dimensional vector spaces over $\mathbb{C}$ for all $i \geq 0$, then for any irreducible representation π of $\mathrm{SO}(V)$ and any irreducible admissible representation σ of $\mathrm{SO}(W)$, $\mathrm{Ext}^i_{\mathrm{SO}(W)}[\pi, \sigma]$ are finite dimensional vector spaces over $\mathbb{C}$ for all $i \geq 0$.*

Proof. By Proposition 2.3, if π is a supercuspidal representation of $\mathrm{SO}(V)$, its restriction to $\mathrm{SO}(W)$ is a projective object in $\mathfrak{R}(SO(W))$. Therefore $\mathrm{Ext}^i_{SO(W)}[\pi, \sigma]$ are zero for $i > 0$, and $\mathrm{Ext}^0_{SO(W)}[\pi, \sigma] = \mathrm{Hom}_{SO(W)}[\pi, \sigma]$ is finite dimensional.

Assume now that π is not a supercuspidal representation, and write π as a quotient of a principal series representation induced from a representation of a maximal parabolic subgroup of $\mathrm{SO}(V)$. We thus have an exact sequence,

$$0 \to \lambda \to Ps \to \pi \to 0.$$

This gives rise to a long exact sequence,

$$0 \to \mathrm{Hom}_{\mathrm{SO}(W)}[\pi, \sigma] \to \mathrm{Hom}_{\mathrm{SO}(W)}[Ps, \sigma] \to \mathrm{Hom}_{\mathrm{SO}(W)}[\lambda, \sigma] \to$$
$$\to \mathrm{Ext}^1_{\mathrm{SO}(W)}[\pi, \sigma] \to \mathrm{Ext}^1_{\mathrm{SO}(W)}[Ps, \sigma] \to \mathrm{Ext}^1_{\mathrm{SO}(W)}[\lambda, \sigma] \to \mathrm{Ext}^2_{\mathrm{SO}(W)}[\pi, \sigma] \to \cdots$$

Since we know that all Hom spaces in the above exact sequence are finite dimensional, and also $\mathrm{Ext}^1_{\mathrm{SO}(W)}[Ps, \sigma]$ is given to be finite dimensional, we get the finite dimensionality of $\mathrm{Ext}^1_{\mathrm{SO}(W)}[\pi, \sigma]$ for any irreducible representation π of $\mathrm{SO}(V)$. This implies finite dimensionality of $\mathrm{Ext}^1_{\mathrm{SO}(W)}[\pi, \sigma]$ for any finite length representation π of $\mathrm{SO}(V)$. Armed with this finite dimensionality of $\mathrm{Ext}^1_{\mathrm{SO}(W)}[\pi, \sigma]$ for any finite length representation π of $\mathrm{SO}(V)$, and with the knowledge that $\mathrm{Ext}^2_{\mathrm{SO}(W)}[Ps, \sigma]$ is given to be finite dimensional, we get the finite dimensionality of $\mathrm{Ext}^2_{\mathrm{SO}(W)}[\pi, \sigma]$ for any irreducible representation π of $\mathrm{SO}(V)$, and similarly we get the finite dimensionality of $\mathrm{Ext}^i_{\mathrm{SO}(W)}[\pi, \sigma]$ for any irreducible representation π of $\mathrm{SO}(V)$, and any $i \geq 0$. □

The proof of Theorem 6.1 uses Theorem 3.1 (Kunneth theorem) for representations of $\mathrm{GL}_m(F) \times G'_k$. Since for any two irreducible representations V, V' of $\mathrm{GL}_m(F)$, $\mathrm{EP}_{\mathrm{GL}_m(F)}[V, V'] = 0$ unless $m = 0$ cf. Lemma 4.1, we obtain the following corollary of the proof of the theorem.

Corollary 6.2. *For a principal series representation $\pi = \pi_0 \rtimes \sigma_0$ of $\mathrm{SO}(V)$ where σ_0 is an admissible representation of $\mathrm{SO}(W)$, and π' is an admissible representation of $\mathrm{SO}(V')$ where $W \subset V' \subset V$ with V' a nondegenerate codimension 1 subspace of the quadratic space V with $\dim(V') - \dim(W) = 2m - 1$,*

$$\mathrm{EP}_{\mathrm{SO}(V')}[\pi, \pi'] = \mathrm{EP}_{\mathrm{Bes}(V', W)}[\pi', \sigma_0] \cdot \dim \mathrm{Wh}(\pi_0).$$

Definition: A finite length representation π of a classical group will be called a *full principal series* if it is irreducible and supercuspidal, or is of the form $\pi = \pi_0 \rtimes \sigma_0$ with both π_0 and σ_0 irreducible, and σ_0 supercuspidal.

The following corollary is a consequence of the previous corollary together with the fact that if σ is a cuspidal representation of $\mathrm{SO}(W)$, then $\sigma \otimes \psi$ is an injective module for $\mathrm{Bes}(V, W)$.

Corollary 6.3. *Let π be a finite length representation of $\mathrm{SO}(V)$, and π' of $\mathrm{SO}(V')$ where $V' \subset V$ is a nondegenerate codimension 1 subspace of the quadratic space V. Assume that π is a full principal series, and π' is an irreducible representation of $\mathrm{SO}(V')$. Then, $\mathrm{EP}_{\mathrm{SO}(V')}[\pi, \pi']$ is either 0 or 1. If $\pi = \pi_0 \rtimes \sigma_0$ of $\mathrm{SO}(V)$ where σ_0 is an admissible representation of $\mathrm{SO}(W)$ with $\dim W \leq 1$, $\mathrm{EP}_{\mathrm{SO}(V')}[\pi, \pi'] = \dim \mathrm{Wh}(\pi) \cdot \dim \mathrm{Wh}(\pi')$ (if $\dim W = 1$, $\mathrm{Wh}(\pi')$ is for a particular character of a maximal unipotent subgroup of $\mathrm{SO}(V')$).*

Remark 6.1. In the previous corollary, we see a large number of cases when the Euler-Poincaré characteristic is 0 or 1. Is there a multiplicity one result for EP, or for Ext^i?

7 An integral formula of Waldspurger, and a conjecture on E-P

In this section we review an integral formula of Waldspurger which we then propose to be the integral formula for the Euler-Poincaré pairing for $\mathrm{EP}_{\mathrm{Bes}(V,W)}[\sigma, \sigma']$ for σ any finite length representation of $\mathrm{SO}(V)$, and σ' any finite length representation of $\mathrm{SO}(W)$, where V and W are quadratic spaces over F with $V = X + D + W + Y$ with W a quadratic subspace of V of codimension $2k + 1$ with X and Y totally isotropic subspaces of V in duality with each other under the underlying bilinear form, and D an anisotropic line in V. Let $Z = X + Y$.

Let $\underline{\mathfrak{T}}$ denote the set of elliptic tori T in $\mathrm{SO}(W)$ such that there exist quadratic subspaces W_T, W_T' of W such that:

1. $W = W_T \oplus W_T'$, and $V = W_T \oplus W_T' \oplus D \oplus Z$.
2. $\dim(W_T)$ is even, and $\mathrm{SO}(W_T')$ and $\mathrm{SO}(W_T' \oplus D \oplus Z)$ are quasi-split.
3. T is a maximal (elliptic) torus in $\mathrm{SO}(W_T)$.

Clearly the group $\mathrm{SO}(W)$ operates on $\underline{\mathfrak{T}}$. Let $\mathfrak{T}$ denote a set of orbits for this action of $\mathrm{SO}(W)$ on $\underline{\mathfrak{T}}$. For our purposes we note the most important elliptic torus $T = \langle e \rangle$ corresponding to $W_T = 0$.

For σ an admissible representation of $\mathrm{SO}(V)$ of finite length, define a function $c_\sigma(t)$ for regular elements of a torus T belonging to $\underline{\mathfrak{T}}$ by the germ expansion of the character $\theta_\sigma(t)$ of σ on the centralizer of t in the Lie algebra of $\mathrm{SO}(V)$, and picking out 'the' leading term. The semi-simple part of the centralizer of t in the Lie algebra of V is the Lie algebra of $\mathrm{SO}(W_T' \oplus D \oplus Z)$ which, if $W_T' \oplus D \oplus Z$ has odd dimension, has a unique conjugacy class of regular nilpotent element, but if $W_T' \oplus D \oplus Z$ has even dimension, then although there are several regular nilpotent conjugacy classes, there is one which is 'relevant', and is what is used to define $c_\sigma(t)$. Similarly, for σ' an admissible representation of $\mathrm{SO}(W)$ of finite length, one defines a function $c_{\sigma'}(t)$ for regular elements of a torus T belonging to $\underline{\mathfrak{T}}$ by the germ expansion of the character $\theta_{\sigma'}(t)$ of σ'.

Define a function Δ_T on an elliptic torus T belonging to $\underline{\mathfrak{T}}$ with $W = W_T \oplus W_T'$, by $\Delta(t) = |\det(1-t)|_{W_T}|_F$, and let D^H denote the discriminant function on $H(F)$. For a torus T in H, define the Weyl group $W(H,T)$ by the usual normalizer divided by the centralizer: $W(H,T) = N_{H(F)}(T)/Z_{H(F)}(T)$.

The following theorem is proved by Waldspurger in J.-L.Waldspurger [2010] and J.-L.Waldspurger [2012b].

Theorem 7.1. *Let $V = X + D + W + Y$ be a quadratic space over the non-archimedean local field F with W a quadratic subspace of codimension $2k+1$ as above. Then for any irreducible admissible representation σ of $\mathrm{SO}(V)$ and irreducible admissible representation σ' of $\mathrm{SO}(W)$,*

$$\sum_{T \in \mathfrak{T}} |W(H,T)|^{-1} \int_{T(F)} c_\sigma(t) c_{\sigma'}(t) D^H(t) \Delta^k(t) dt,$$

is a finite sum of absolutely convergent integrals. (The Haar measure on $T(F)$ is normalized to have volume 1.) If either σ is a supercuspidal representation of $\mathrm{SO}(V)$*, and σ' is arbitrary irreducible admissible representation of* $\mathrm{SO}(W)$*, or both σ and σ' are tempered representations, then*

$$\dim \mathrm{Hom}_{\mathrm{Bes}(V,W)}[\sigma, \sigma'] = \sum_{T \in \mathcal{T}} |W(H,T)|^{-1} \int_{T(F)} c_\sigma(t) c_{\sigma'}(t) D^H(t) \Delta^k(t) dt.$$

Given this theorem of Waldspurger, it is most natural to propose the following conjecture on Euler-Poincaré pairing.

Conjecture 7.1. *Let $V = X + D + W + Y$ be a quadratic space over the non-archimedean local field F with W a quadratic subspace of V of codimension $2k+1$ as before. Then for any irreducible admissible representation σ of* $\mathrm{SO}(V)$ *and irreducible admissible representation σ' of* $\mathrm{SO}(W)$*,*

1.

$$\begin{aligned} \mathrm{EP}_{\mathrm{Bes}(V,W)}[\sigma, \sigma'] &= \sum_i (-1)^i \dim \mathrm{Ext}^i_{\mathrm{Bes}(V,W)}[\sigma, \sigma'] \\ &= \sum_{T \in \mathcal{T}} |W(H,T)|^{-1} \int_{T(F)} c_\sigma(t) c_{\sigma'}(t) D^H(t) \Delta^k(t) dt. \end{aligned}$$

2. *If σ and σ' are irreducible tempered representations, then* $\mathrm{Ext}^i_{\mathrm{Bes}(V,W)}[\sigma, \sigma'] = 0$ *for $i > 0$.*

Remark 7.1. Note that a supercuspidal representation of $\mathrm{SO}(V)$ is a projective object in the category of smooth representations of $\mathrm{SO}(V)$, and hence by Proposition 2.4, it remains a projective object in the category of smooth representations of $\mathrm{SO}(W) \cdot U$. Therefore if σ or σ' is supercuspidal, $\mathrm{Ext}^i_{\mathrm{Bes}(V,W)}[\sigma, \sigma'] = 0$ for $i > 0$. (We note that for a supercuspidal representation σ' of $\mathrm{SO}(W)$, the representation $\sigma' \otimes \psi$ is an injective module in the category of smooth representations of $\mathrm{SO}(W) \cdot U$.) Thus Waldspurger's theorem is equivalent to the conjectural statement on Euler-Poincaré characteristic if σ or σ' is supercuspidal (except that it is not proved if σ' is supercuspidal, but σ is arbitrary). Part 2 of the conjecture is there as the simplest possible explanation of Waldspurger's theorem for tempered representations!

Example 7.1. Assume that either $G = \mathrm{SO}_{n+1}(F)$ is a split group, and σ is induced from a character of a Borel subgroup of G, or $H = \mathrm{SO}_n(F)$ is a split group and σ' is induced from a character of a Borel subgroup of H. Then the conjectural formula on Euler-Poincaré becomes $\mathrm{EP}[\sigma, \sigma'] = 1$ which is a consequence of Corollary 6.3.

Remark 7.2. We consider the Waldspurger integral formula as some kind of Riemann-Roch theorem. Recall that for X a smooth projective variety with Todd class T_X, and for any coherent sheaf $\mathfrak{F}$ on X with Chern class $c(\mathfrak{F})$, one has,

$$EP(X, \mathfrak{F}) = \sum_i (-1)^i \dim H^i(X, \mathfrak{F}) = \sum_i (-1)^i \dim \mathrm{Ext}^i(\mathcal{O}_X, \mathfrak{F}) = \int_X (T_X \cdot c(\mathfrak{F})).$$

In our case, $EP[\pi_1, \pi_2] = \sum_i (-1)^i \dim \mathrm{Ext}^i_H[\pi_1, \pi_2]$ is conjecturally expressed as

$$EP[\pi_1, \pi_2] = \int_X T_X \cdot c(\pi_1, \pi_2),$$

where X is a certain set of elliptic tori in H, T_X is a function on this set of elliptic tori, and $c(\pi_1, \pi_2)$ is a function on these elliptic tori defined in terms of the germ expansion of π_1 and π_2.

8 The Schneider-Stuhler duality theorem

The following theorem is a mild generalization of a duality theorem of Schneider and Stuhler due to M.Nori and Prasad [2017]; it turns questions on $\mathrm{Ext}^i[\pi_1, \pi_2]$ to $\mathrm{Ext}^j[\pi_2, \pi_1]$, and therefore is of central importance to our theme in this paper.

Theorem 8.1. *Let G be a reductive p-adic group, and π an irreducible, admissible representation of G. Let $d(\pi)$ be the largest integer $i \geq 0$ such that there is an irreducible, admissible representation π' of G with $\mathrm{Ext}^i_G[\pi, \pi']$ nonzero. Then,*

1. *There is a unique irreducible representation π' of G with $\mathrm{Ext}^{d(\pi)}_G[\pi, \pi'] \neq 0$.*
2. *The representation π' in (1) is nothing but $D(\pi)$ where $D(\pi)$ is the Aubert-Zelevinsky involution of π, and $d(\pi)$ is the split rank of the Levi subgroup M of G which carries the cuspidal support of π.*
3. $\mathrm{Ext}^{d(\pi)}_G[\pi, D(\pi)] \cong \mathbb{C}$.
4. *For any smooth representation π' of G, the bilinear pairing*

$$(*) \quad \mathrm{Ext}^i_G[\pi, \pi'] \times \mathrm{Ext}^j_G[\pi', D(\pi)] \to \mathrm{Ext}^{i+j=d(\pi)}_G[\pi, D(\pi)] \cong \mathbb{C},$$

is nondegenerate in the sense that if $\pi' = \varinjlim \pi'_n$ of finitely generated G-submodules π'_n, then $\mathrm{Ext}^i_G[\pi, \pi'] = \varinjlim \mathrm{Ext}^i_G[\pi, \pi'_n]$, a direct limit of finite dimensional vector spaces over $\mathbb{C}$, and $\mathrm{Ext}^j_G[\pi', D(\pi)] = \varprojlim \mathrm{Ext}^j_G[\pi'_n, D(\pi)]$, an inverse limit *of finite dimensional vector spaces over $\mathbb{C}$, and the pairing in $(*)$ is the direct limit of perfect pairings on these finite dimensional spaces:*

$$\mathrm{Ext}_G^i[\pi, \pi_n'] \times \mathrm{Ext}_G^j[\pi_n', D(\pi)] \to \mathrm{Ext}_G^{i+j=d(\pi)}[\pi, D(\pi)] \cong \mathbb{C}.$$

(Observe that a compatible family of perfect pairings on finite dimensional vector spaces $B_n : V_n \times W_n \to \mathbb{C}$ with V_n part of an inductive system, and W_n part of a projective system, gives rise to a natural pairing $B : \varinjlim V_n \times \varprojlim W_n \to \mathbb{C}$ such that the associated homomorphism from $(\varinjlim V_n)^\star$ to $\varprojlim W_n$ is an isomorphism.

As an example, the following proposition giving complete classification of irreducible submodules π of the tensor product $\pi_1 \otimes \pi_2$ of two (irreducible, infinite dimensional) representations π_1, π_2 of $\mathrm{GL}_2(F)$ with the product of their central characters trivial, is essentially a translation of vanishing of $\mathrm{Ext}^1_{PGL_2(F)}[\pi_1 \otimes \pi_2, \pi_3]$ by this duality theorem. The vanishing itself follows because $\mathrm{Ext}^2_{PGL_2(F)}[\pi_1 \otimes \pi_2, \pi_3] = 0$ by Proposition 2.9, and both EP and Hom spaces have the same dimension.

Proposition 8.1. *Let π_1, π_2 be two irreducible admissible infinite dimensional representations of $\mathrm{GL}_2(F)$ with product of their central characters trivial. Then the following is the complete list of irreducible sub-representations π of $\pi_1 \otimes \pi_2$ as $\mathrm{PGL}_2(F)$-modules.*

1. *π is a supercuspidal representation of $\mathrm{PGL}_2(F)$, and appears as a quotient of $\pi_1 \otimes \pi_2$.*
2. *π is a twist of the Steinberg representation, which we assume by absorbing the twist in π_1 or π_2 to be the Steinberg representation St of $\mathrm{PGL}_2(F)$. Then St is a submodule of $\pi_1 \otimes \pi_2$ if and only if π_1, π_2 are both irreducible principal series representations, and $\pi_1 \cong \pi_2^\vee$.*

9 Geometrization of Ext groups

A natural way to construct exact sequences in representation theory is via the Bernstein-Zelevinsky exact sequence arising from the inclusion of an open set $X - Y$ in a topological space X equipped with an ℓ-sheaf $\mathfrak{F}$, with Y a closed subspace of X, giving rise to

$$0 \to \mathcal{S}(X - Y, \mathfrak{F}) \to \mathcal{S}(X, \mathfrak{F}) \to \mathcal{S}(Y, \mathfrak{F}) \to 0.$$

Observe that in this exact sequence, the larger space $\mathcal{S}(X - Y, \mathfrak{F})$ arises as a subspace, whereas the smaller space $\mathcal{S}(Y, \mathfrak{F})$ arises as a quotient of $\mathcal{S}(X, \mathfrak{F})$. Assuming that a group G operates on the space X, preserving the closed subspace Y, as well as the sheaf $\mathfrak{F}$, then this exact sequence gives rise to an element of $\mathrm{Hom}_G[\mathcal{S}(X, \mathfrak{F}), \mathcal{S}(Y, \mathfrak{F})]$, as well as an element of $\mathrm{Ext}^1_G[\mathcal{S}(Y,\mathfrak{F}),\mathcal{S}(X-Y,\mathfrak{F})]$. Note that the Hom is from a larger space $\mathcal{S}(X,\mathfrak{F})$ to a smaller space $\mathcal{S}(Y,\mathfrak{F})$, whereas the Ext is between a smaller space and a larger space.

Similarly, if X_2, X_1 are closed subsets of an ℓ-space X with $X_2 \subset X_1 \subset X$, and endowed with an ℓ-sheaf $\mathfrak{F}$, we have exact sequences,

$$0 \to \mathcal{S}(X-X_1,\mathfrak{F}) \to \mathcal{S}(X-X_2,\mathfrak{F}) \to \mathcal{S}(X_1-X_2,\mathfrak{F}) \to 0,$$

$$0 \to \mathcal{S}(X_1-X_2,\mathfrak{F}) \to \mathcal{S}(X_1,\mathfrak{F}) \to \mathcal{S}(X_2,\mathfrak{F}) \to 0,$$

which can be spliced together to give rise to the exact sequence,

$$0 \to \mathcal{S}(X-X_1,\mathfrak{F}) \to \mathcal{S}(X-X_2,\mathfrak{F}) \to \mathcal{S}(X_1,\mathfrak{F}) \to \mathcal{S}(X_2,\mathfrak{F}) \to 0,$$

which gives an element of $\mathrm{Ext}^2_G[\mathcal{S}(X_2,\mathfrak{F}),\mathcal{S}(X-X_1,\mathfrak{F})]$; so as the representation $\pi_2 = \mathcal{S}(X_2,\mathfrak{F})$ becomes smaller and smaller compared to $\pi_1 = \mathcal{S}(X-X_1,\mathfrak{F})$ (as the space X_2 is 'two step smaller' than X), it may be expected to contribute to higher and higher Ext groups $\mathrm{Ext}^i_G[\pi_2,\pi_1]$.

Various examples around the present work suggest that homomorphisms between representations, or extensions between them correspond to some geometric spaces (and ℓ-sheaves on them) as above, in particular, a typical homomorphism is from a larger space to smaller ones, whereas a typical Ext is the other way around!

Although most geometric spaces have algebraic geometric origin, that is not necessarily the case when thinking about the Bernstein-Zelevinsky exact sequence. For instance, one can use the action of G on its Bruhat-Tits building and its various compactifications. If X is the tree associated to $\mathrm{PGL}_2(F)$, then one knows that there is a compactification $\overline{X}$ of X on which $\mathrm{PGL}_2(F)$ continues to act with $\overline{X} - X = \mathbb{P}^1(F)$, a closed subset of $\overline{X}$. The zero-skeleton X^0 of X together with $\overline{X} - X$ is a compact topological space, call it $\overline{X}_0$, with an action of $\mathrm{PGL}_2(F)$ with two orbits: X^0 which is the open orbit, and $\mathbb{P}^1(F)$ which is the closed orbit.

An unramified character χ of B gives rise to a sheaf, say $\mathbb{C}_\chi$, on $\mathbb{P}^1(F)$, which can be extended to a $\mathrm{PGL}_2(F)$-equivariant sheaf on $\overline{X}_0$ by making it $\mathrm{ind}_{\mathrm{PGL}_2(\mathcal{O}_F)}^{\mathrm{PGL}_2(F)}\mathbb{C}$ on X^0. Call the extended sheaf on $\overline{X}_0$ also as $\mathbb{C}_\chi$; note that the restriction of $\mathbb{C}_\chi$ to X^0 is the constant sheaf $\mathbb{C}$. Thus we have an exact sequence,

$$0 \to \mathcal{S}(X^0,\mathbb{C}) \to \mathcal{S}(\overline{X}_0,\mathbb{C}_\chi) \to \mathcal{S}(\mathbb{P}^1(F),\mathbb{C}_\chi) \to 0.$$

Since $\mathrm{PGL}_2(F)$ acts transitively on the zero-skeleton X^0 with stabilizer $\mathrm{PGL}_2(\mathcal{O}_F)$, we have $\mathcal{S}(X^0,\mathbb{C}) \cong \mathrm{ind}_{\mathrm{PGL}_2(\mathcal{O}_F)}^{\mathrm{PGL}_2(F)}\mathbb{C}$. We have thus constructed an element of

$$\mathrm{Ext}^1_{\mathrm{PGL}_2(F)}[\mathrm{Ind}_B^{\mathrm{PGL}_2(F)}\chi, \mathrm{ind}_{\mathrm{PGL}_2(\mathcal{O}_F)}^{\mathrm{PGL}_2(F)}\mathbb{C}]$$

It may be hoped that many extensions which representation theory offers will be matched by geometric action of G on topological spaces with finitely many G-orbits coming either from algebraic geometric spaces, or from the Bruhat-Tits building and its compactifications; there is then also the issue of proving that geometric actions do give non-trivial extensions!

We end with a precise question but before that we need to make a definition. In what follows, groups and spaces are what are called ℓ-groups and ℓ-spaces in Bernstein and Zelevinsky [1976] and Bernstein [1992]; we will also use the notion of an ℓ-sheaf from these references which we recall is a sheaf, say $\mathcal{F}$, over X of vector spaces over $\mathbb{C}$ with the space of compactly supported global sections $\mathcal{S}(X, \mathcal{F}) = \mathcal{F}_c(X)$, a module over $\mathcal{S}(X)$, which is nondegenerate in the sense that $\mathcal{S}(X) \cdot \mathcal{F}_c(X) = \mathcal{F}_c(X)$; the functor $\mathcal{F} \to \mathcal{F}_c(X)$ gives an equivalence between ℓ-sheaves on X and nondegenerate modules over $\mathcal{S}(X)$.

Definition: A complex representation V of $G = G(F)$ is said to be of geometric origin if

1. there is a G-space X_V with finitely many G-orbits,
2. a G-equivariant sheaf $\mathcal{F}_V$ on X_V,

such that on each G-orbit $Y \subset X_V$ of the form G/H_Y, $\mathcal{F}_V|_Y$ is the equivariant sheaf associated to a finite dimensional representation W_Y of H_Y, and $V \cong \mathcal{S}(X_V, \mathcal{F}_V)$ (cf. §1.16 of Bernstein and Zelevinsky [1976] for the definition of the restriction of an ℓ-sheaf to a locally closed subset such as Y).

Example 9.1. Parabolic induction and Jacquet functor take representations of geometric origin to representations of geometric origin. It is expected that all supercuspidal representations are of geometric origin (proved for $\mathrm{GL}_n(F)$ and classical groups in odd residue characteristic). In understanding the class of representations of geometric origin given by the action of a group G on a space X, one difficulty seems to be to glue vector bundles on various orbits to an ℓ-sheaf on X.

Question 9.1. Suppose that we have two complex smooth representations V_1 and V_2 of $G = G(F)$ of geometric origin with $\mathrm{Ext}^1_G[V_1, V_2] \neq 0$ with $V \in \mathrm{Ext}^1_G[V_1, V_2]$ represented by the extension

$$0 \to V_2 \to V \to V_1 \to 0.$$

Then is V of geometric origin?

Remark 9.1. A representation of *geometric origin* comes equipped with considerable additional data as in the Bernstein–Zelevinsky exact sequence, which may be important to refine the question above.

References

A. Aizenbud, D. Gourevitch, S. Rallis, and G. Schiffmann (2010). "Multiplicity One Theorems". *Annals of Mathematics* 172, no. 2, pp. 1407–1434 (cit. on p. 1386).

A. Aizenbud and E. Sayag (2017). "Homological multiplicities in representation theory of p-adic groups". arXiv: 1709.09886 (cit. on p. 1386).

I. N. Bernstein (1992). "Representations of p-adic groups". *Written by Karl E. Rumelhart, Harvard University* (cit. on pp. 1390, 1411).

I. N. Bernstein and A. V. Zelevinsky (1976). "Representations of the group $\mathrm{GL}(n, F)$ where F is a non-Archimdean local field". *Russian Math Surveys* 31.3, pp. 1–68 (cit. on pp. 1395, 1411).

R. Beuzart-Plessis (2014). "Expression d'un facteur epsilon de paire par une formule intégrale". *Canad. J Math* 66, No. 5, pp. 993–1049 (cit. on p. 1386).

– (2015). "Endoscopie et conjecture raffinée de Gan-Gross-Prasad pour les groupes unitaires". *Compositio Math* 151, No. 7, pp. 1309–1371 (cit. on p. 1386).

– (2016). "La conjecture locale de Gross-Prasad pour les reprśentations tempérées des groupes unitaires". *Mém. Soc. Math. Fr. (N.S.)* 149, vii+191 pages (cit. on p. 1386).

K. Y. Chan and G. Savin (2017). "Bernstein-Zelevinsky derivatives, branching rules and Hecke algebras". arXiv: 1605.05130 (cit. on p. 1397).

W. T. Gan, B. Gross, and D. Prasad (2012). "Symplectic local root numbers, central critical L-values, and restriction problems in the representation theory of classical groups". *Astérisque* 346, pp. 1–109 (cit. on pp. 1386, 1401).

J.-L.Waldspurger (2010). " Une Formule intégrale reliée à la conjecture locale de Gross-Prasad". *Compositio Math.* 146, no. 5, 1180-1290. (Cit. on pp. 1386, 1387, 1406).

– (2012a). " La conjecture locale de Gross-Prasad pour les représentations tempérées des groupes spéciaux orthogonaux". *Astérisque* 347, pp. 103–165 (cit. on p. 1386).

– (2012b). " Une Formule intégrale reliée à la conjecture locale de Gross-Prasad, 2ème partie: extension aux représentations tempérées". *Astérisque* 346, pp. 171–312 (cit. on pp. 1386, 1387, 1406).

T. Kobayashi (1994). "Discrete decomposability of the restriction of $A_{\mathfrak{q}}(\lambda)$ with respect to reductive subgroups and its applications." *Invent. Math.* 117, no. 2, pp. 181–205 (cit. on p. 1387).

M.Nori and D. Prasad (2017). "On a duality theorem of Schneider-Stuhler". arXiv: 1711.01908 (cit. on pp. 1398, 1408).

C. Moeglin and J.-L.Waldspurger (2012). " La conjecture locale de Gross-Prasad pour les groupes spéciaux orthogonaux: le cas général". *Astérisque* 347, pp. 167–216 (cit. on pp. 1386, 1402, 1403).

P.Schneider and U.Stuhler (1997). "Representation theory and sheaves on the Bruhat–Tits building". *Inst. Hautes Études Sci. Publ. Math.* 85, pp. 97–191 (cit. on pp. 1387, 1389, 1392, 1394, 1398).

D. Prasad (1990). "Trilinear forms for representations of $\mathrm{GL}(2)$ and local ϵ-factors". *Compositio Math* 75.1, pp. 1–46 (cit. on p. 1387).

– (1993). "On the decomposition of a representation of $\mathrm{GL}(3)$ restricted to $\mathrm{GL}(2)$ over a p-adic field". *Duke Math. J.* 69.1, pp. 167–177 (cit. on pp. 1394, 1397).

A. Raghuram (2007). "A Kunneth theorem for p-adic groups". *Canad. Math. Bull.* 50.3, pp. 440–446 (cit. on p. 1393).

B. Sun and C. Zhu (2012). "Multiplicity one theorems: the archimedean case". *Annals of Mathematics* 175, No 1, 23-44. (Cit. on p. 1386).

Received 2017-12-05.

Dipendra Prasad
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay - 400 005
India
Laboratory of Modern Algebra and Applications
Saint-Petersburg State University Russia
prasad.dipendra@gmail.com
dprasad@math.tifr.res.in

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1411–1442)

KAC POLYNOMIALS AND LIE ALGEBRAS ASSOCIATED TO QUIVERS AND CURVES

Olivier G. Schiffmann

Abstract

We provide an explicit formula for the following enumerative problem: how many (absolutely) indecomposable vector bundles of a given rank r and degree d are there on a smooth projective curve X of genus g defined over a finite field $\mathbb{F}_q$? The answer turns out to only depend on the genus g, the rank r and the Weil numbers of the curve X. We then provide several interpretations of these numbers, either as the Betti numbers or counting polynomial of the moduli space of stable Higgs bundles (of same rank r and degree d) over X, or as the character of some infinite dimensional graded Lie algebra. We also relate this to the (cohomological) Hall algebras of Higgs bundles on curves and to the dimension of the space of absolutely cuspidal functions on X.

1 Kac polynomials for quivers and curves

1.1 Quivers. Let Q be a locally finite quiver with vertex set I and edge set Ω. For any finite field $\mathbb{F}_q$ and any dimension vector $\mathbf{d} \in \mathbb{N}^I$, let $A_{Q,\mathbf{d}}(\mathbb{F}_q)$ be the number of absolutely indecomposable representations of Q over $\mathbb{F}_q$, of dimension $\mathbf{d}$. Kac proved the following beautiful result:

Theorem 1.1 (Kac [1980]). *There exists a (unique) polynomial $A_{Q,\mathbf{d}}(t) \in \mathbb{Z}[t]$ such that for any finite field $\mathbb{F}_q$,*

$$A_{Q,\mathbf{d}}(\mathbb{F}_q) = A_{Q,\mathbf{d}}(q).$$

Moreover $A_{Q,\mathbf{d}}(t)$ is independent of the orientation of Q and is monic of degree $1 - \langle \mathbf{d}, \mathbf{d} \rangle$.

Here $\langle\ ,\ \rangle$ is the Euler form, see Section 2.1 The fact that the number of (absolutely) indecomposable $\mathbb{F}_q$-representations behaves polynomially in q is very remarkable: the absolutely indecomposable representations only form a *constructible* substack of the stack

MSC2010: primary 17B65; secondary 14D20.

$\mathfrak{M}_{Q,\mathbf{d}}$ of representations of Q of dimension $\mathbf{d}$ and we count them here *up to isomorphism*, i.e. *without* the usual orbifold measure. Let us briefly sketch the idea of a proof. Standard Galois cohomology arguments ensure that it is enough to prove that the number of indecomposable $\mathbb{F}_q$-representations is given by a polynomial $I_{Q,\mathbf{d}}$ in q (this polynomial is not as well behaved as or as interesting as $A_{Q,\mathbf{d}}$). Next, by the Krull-Schmidt theorem, it is enough to prove that the number of *all* $\mathbb{F}_q$-representations of dimension $\mathbf{d}$ is itself given by a polynomial in $\mathbb{F}_q$. This amounts to computing the (orbifold) volume of the *inertia* stack $\mathfrak{I}\mathfrak{M}_{Q,\mathbf{d}}$ of $\mathfrak{M}_{Q,\mathbf{d}}$; performing a unipotent reduction in this context, we are left to computing the volume of the stack $\mathfrak{N}il_{Q,\mathbf{d}}$ parametrizing pairs (M, ϕ) with M a representation of dimension $\mathbf{d}$ and $\phi \in \mathrm{End}(M)$ being nilpotent. Finally, we use a Jordan stratification of $\mathfrak{N}il_{Q,\mathbf{d}}$ and easily compute the volume of each strata in terms of the volumes of the stacks $\mathfrak{M}_{Q,\mathbf{d}'}$ for all $\mathbf{d}'$. This actually yields an explicit formula for $A_{Q,\mathbf{d}}(t)$ (or $I_{Q,\mathbf{d}}(t)$), see Hua [2000]. We stress that beyond the case of a few quivers (i.e. those of finite or affine Dynkin type) it is unimaginable to classify and construct all indecomposable representations; nevertheless, the above theorem says that we can *count* them.

The positivity of $A_{Q,\mathbf{d}}(t)$ was only recently[1]:

Theorem 1.2 (Hausel, Letellier, and Rodriguez-Villegas [2013]). *For any Q and any $\mathbf{d}$ we have $A_{Q,\mathbf{d}}(t) \in \mathbb{N}[t]$.*

The desire to understand the meaning of the Kac polynomials $A_{Q,\mathbf{d}}(t)$, for instance as the dimension of certain natural graded vector spaces (such as the cohomology of some algebraic variety), has been a tremendous source of inspiration in geometric representation theory; it has lead to the development over the years of a beautiful and very rich theory relating various moduli spaces of representations of quivers to Lie algebras and quantum groups, yielding geometric constructions of a host of important objects in representation theory (such as canonical and crystal bases, Yangians, quantum affine algebras, their highest weight or finite-dimensional representations, etc.). We will present here some recent developments in this area, in particular in relation to the structure and representation theory of Kac-Moody algebras and *graded Borcherds algebras*.

1.2 Curves. Now let X be a smooth projective curve defined over some finite field $\mathbb{F}_q$. For any $r \geq 0, d \in \mathbb{Z}$ let $A_{r,d}(X)$ be the number of isomorphism classes of absolutely indecomposable coherent sheaves $\mathfrak{F}$ over X of rank r and degree d. It turns out that this number may also be computed explicitly and depends 'polynomially' on X. In order to make sense of this, we need to introduce a few notations. The action of the geometric Frobenius Fr_X on the l-adic cohomology group $H^1_{et}(X_{\overline{\mathbb{F}_q}}, \overline{\mathbb{Q}_l})$ is semisimple with

[1] the special case of an indivisible dimension vector was proved earlier by Crawley-Boevey and Van den Bergh [2004]

eigenvalues $\{\sigma_1, \ldots, \sigma_{2g}\}$ which may be ordered so that $\sigma_{2i-1}\sigma_{2i} = q$ for all i. Moreover, Fr_X belongs to the general symplectic group $GSp(H^1_{et}(X_{\overline{\mathbb{F}_q}}, \overline{\mathbb{Q}_l}))$ relative to the intersection form on $H^1_{et}(X_{\overline{\mathbb{F}_q}}, \overline{\mathbb{Q}_l})$. We may canonically identify the character ring of $GSp(H^1_{et}(X_{\overline{\mathbb{F}_q}}, \overline{\mathbb{Q}_l}))$ with $R_g = \overline{\mathbb{Q}_l}[T_g]^{W_g}$ where

$$T_g = \{(\eta_1, \ldots, \eta_{2g}) \in \mathbb{G}_m^{2g} \mid \eta_{2i-1}\eta_{2i} = \eta_{2j-1}\eta_{2j}\ \forall\, i, j\}, \qquad W_g = (\mathfrak{S}_2)^g \rtimes \mathfrak{S}_g$$

are the maximal torus, resp. Weyl group of $GSp(2g, \overline{\mathbb{Q}_l})$. We can evaluate any element $f \in R_g$ on Fr_X for any smooth projective curve X of genus g, defined over some[2] finite field $\mathbb{F}_q$; concretely, we have

$$f(Fr_X) = f(\sigma_1, \sigma_2, \ldots, \sigma_{2g}).$$

We will say that a quantity depending on X is *polynomial* if it is the evaluation of some element $f \in R_g$. For instance the size of the field $\mathbb{F}_q$ of definition of X is polynomial in X (take $f = \eta_{2i-1}\eta_{2i}$ for any i), as is the number $|X(\mathbb{F}_q)|$ of $\mathbb{F}_q$-rational points of X (take $f = 1 - \sum_i \eta_i + q$); this is also true of symmetric powers $S^l X$ of X. We extend the above definition to the case $g = 0$ by setting $R_g = \overline{\mathbb{Q}_l}[q^{\pm 1}]$.

Theorem 1.3 (Schiffmann [2016]). *For any fixed g, r, d there exists a (unique) polynomial $A_{g,r,d} \in R_g$ such that for any smooth projective curve X of genus g defined over a finite field, the number of absolutely indecomposable coherent sheaves on X of rank r and degree d is equal to $A_{g,r,d}(Fr_X)$. Moreover, $A_{g,r,d}$ is monic of leading term $q^{1+(g-1)r^2}$.*

The finiteness of the number of absolutely indecomposable coherent sheaves of fixed rank and degree is a consequence of Harder-Narasimhan reduction theory; it suffices to observe that any sufficiently unstable coherent sheaf is decomposable as its Harder-Narasimhan filtration splits in some place. Let us say a few words about the proof of Theorem 1.3: like for quivers, standard Galois cohomology arguments combined with the Krull-Schmidt theorem reduce the problem to counting *all* isomorphism classes of coherent sheaves of rank r and degree d; unfortunately, this number is infinite as soon as $r > 0$ hence it is necessary to introduce a suitable truncation of the category $\mathrm{Coh}(X)$; there are several possibilities here, one of them being to consider the category $\mathrm{Coh}^{\geq 0}(X)$ of positive coherent sheaves, i.e. sheaves whose HN factors all have positive degree. One is thus lead to compute the volumes of the inertia stacks $\mathfrak{I}\mathfrak{M}^{\geq 0}_{X,r,d}$ of the stacks $\mathfrak{M}^{\geq 0}_{X,r,d}$ positive sheaves of rank r and degree d on X. Finally, after a unipotent and Jordan reduction similar to the case of quivers, this last computation boils down to the evaluation of the integral of certain Eisenstein series over the truncated stacks $\mathfrak{M}^{\geq 0}_{X,r,d}$; this is performed using a variant of Harder's method for computing the volume of $\mathfrak{M}_{X,r,d}$, see Schiffmann [ibid.]

[2] technically, of characteristic different from l

for details. The proof is constructive and yields an explicit but complicated formula for $A_{g,r,d}$. Hence, just like for quivers, although classifying all indecomposable vector bundles for curves of genus $g > 1$ is a wild problem, it is nevertheless possible to count them. The explicit formula for $A_{g,r,d}$ was later combinatorially very much simplified by Mellit [2017b], who in particular proved the following result:

Theorem 1.4 (Mellit [ibid.]). *The polynomial $A_{g,r,d}$ is independent of d.*

Thanks to this theorem, we may simply write $A_{g,r}$ for $A_{g,r,d}$. Like their quivery cousins, Kac polynomials of curves satisfy some positivity and integrality property, namely

Theorem 1.5 (Schiffmann [2016]). *For any g, r, $A_{g,r} \in Im(\mathbb{N}[-\eta_1, \ldots, -\eta_g]^{W_g} \to R_g)$.*

In other words, $A_{g,r}$ is a W_g-symmetric polynomial in $-\eta_1, \ldots, -\eta_{2g}$ with positive integral coefficients. However, this is *not* the most natural form of positivity:

Conjecture 1.6. *For any g, r there exists a canonical (*non-*virtual) finite-dimensional representation $\mathbb{A}_{g,r}$ of $GSp(2g, \overline{\mathbb{Q}_l})$ such that $A_{g,r} = \tau(ch(\mathbb{A}_{g,r}))$, where $\tau \in Aut(R_g)$ is the involution mapping η_i to $-\eta_i$ for all i.*

The paper Hua [2000] contains many examples of Kac polynomials for quivers. Let us give here the examples of Kac polynomials for curves (of any genus) for $r = 0, 1, 2$; we will write q for $\eta_{2i-1}\eta_{2i}$.

$$A_{g,0} = 1 - \sum_i \eta_i + q, \qquad A_{g,1} = \prod_{i=1}^{2g} (1 - \eta_i),$$

$$A_{g,2} = \prod_{i=1}^{2g}(1-\eta_i) \cdot \left(\frac{\prod_i (1 - q\eta_i)}{(q-1)(q^2-1)} - \frac{\prod_i (1+\eta_i)}{4(1+q)} + \right. \\ \left. + \frac{\prod_i (1-\eta_i)}{2(q-1)} \left[\frac{1}{2} - \frac{1}{q-1} - \sum_i \frac{1}{1-\eta_i} \right] \right).$$

For $r = 0, 1$ we recognize the number of $\mathbb{F}_q$-rational points of X and $Pic^0(X)$ respectively. In particular, we have $\mathbb{A}_{g,0} = \mathbb{C} \oplus \mathbb{V} \oplus det(V)^{1/g}$ and $\mathbb{A}_{g,1} = \Lambda^\bullet V$, where V is the standard $2g$-dimensional representation of $GSp(2g, \overline{\mathbb{Q}_l})$.

Motivated by the analogy with quivers, it is natural to try to seek a representation-theoretic meaning to the Kac polynomials $A_{g,r}$. What is the analog, in this context, of the Kac-Moody algebra associated to a quiver ? Are the Kac polynomials related to the

Poincaré polynomials of some interesting moduli spaces ? Although it is still much less developed than in the context of quivers, we will illustrate this second theme through two examples of applications of the theory of *Hall algebras* of curves: a case of geometric Langlands duality (in the neighborhood of the trivial local system) and the computation of the Poincaré polynomial of the moduli spaces of semistable Higgs bundles on smooth projective curves.

1.3 Quivers vs. Curves. Kac polynomials of quivers and curves are not merely related by an analogy: they are connected through the following observation, which comes by comparing the explicit formulas for $A_{Q,\mathbf{d}}$ and $A_{g,r}$. Let S_g be the quiver with one vertex and g loops.

Proposition 1.7 (Rodriguez-Villegas)**.** *For any r we have $A_{S_g,r}(1) = A_{g,r}(0,\dots,0)$.*

This relation between the constant term of $A_{g,r}$ and the sum of all the coefficients of $A_{S_g,r}$ has a very beautiful conjectural conceptual explanation in terms of the mixed Hodge structure of the (twisted) genus g character variety, we refer the interested reader to Hausel and Rodriguez-Villegas [2008]. We will provide another conceptual explanation in terms of the geometric Langlands duality in Section 5.4.

Remark. There is an entirely similar story for the category of coherent sheaves on a smooth projective curve equipped with a (quasi-)parabolic structure along an effective divisor D, see Schiffmann [2004], Lin [2014], and Mellit [2017a]. Proposition 1.7 still holds in this case, with the quiver S_g being replaced by a quiver with a central vertex carrying g loops, to which are attached finitely many type A branches, one for each point in D.

Plan of the paper. In Sections 2 to 4 we describe various Lie-theoretical (Sections 2.2, 2.5, 3.3, 4.4 and 4.6) or geometric (Sections 4.1 and 4.4 to 4.6) incarnations (some conjectural) of the Kac polynomials for quivers; in particular we advocate the study of a certain *graded* Borcherds algebra $\widetilde{\mathfrak{g}}_Q$ canonically associated to a quiver, which is a graded extension of the usual Kac-Moody algebra $\mathfrak{g}_Q$ attached to Q. From Section 5 onward, we turn our attention to curves and present several algebraic or geometric constructions suggesting the existence of some hidden, combinatorial and Lie-theoretical structures controlling such things as dimensions of spaces of cuspidal functions (Section 6.2), or Poincaré polynomials of moduli spaces of (stable) Higgs bundles on curves (Sections 7.2 and 7.3).

2 Quivers, Kac polynomials and Kac-Moody algebras

We begin with some recollections of some classical results in the theory of quivers and Kac-Moody algebras, including the definitions of Hall algebras and Lusztig nilpotent varieties, and their relations to (quantized) enveloping algebra of Kac-Moody algebras. This is related to the *constant term* $A_{Q,\mathbf{d}}(0)$ of the Kac polynomials $A_{Q,\mathbf{d}}(t)$. The theory is classically developed for quivers without edge loops. The general case, important for applications, is more recent.

2.1 Kac-Moody algebras from quivers. Let $Q = (I, \Omega)$ be a finite quiver without edge loops. For any field $\mathbf{k}$ we denote by $\mathrm{Rep}_{\mathbf{k}}(Q)$ the abelian category of $\mathbf{k}$-representations of Q and for any dimension vector $\mathbf{d} \in \mathbb{N}^I$ we denote by $\mathfrak{M}_{Q,\mathbf{d}}$ the stack of $\mathbf{d}$-dimensional representations of Q. The category $\mathrm{Rep}_{\mathbf{k}}(Q)$ is of global dimension at most one (and exactly one if Ω is non-empty) with finite-dimensional Ext spaces. As a result, the stack $\mathfrak{M}_Q = \bigsqcup_{\mathbf{d}} \mathfrak{M}_{Q,\mathbf{d}}$ is smooth. The first relation to Lie theory arises when considering the Euler forms

(2-1)

$$\langle M, N\rangle = \dim \mathrm{Hom}(M, N) - \dim \mathrm{Ext}^1(M, N), \qquad (M, N) = \langle M, N\rangle + \langle N, M\rangle.$$

Let c_{ij} be the number of arrows in Q from i to j, and let $C = (c_{ij})_{i,j\in I}$ be the adjacency matrix. Set $A = 2Id - C - {}^tC$. The Euler forms $\langle\ ,\ \rangle$ and $(\ ,\ )$ factor through the map

$$\underline{\dim} : K_0(\mathrm{Rep}_{\mathbf{k}} Q) \to \mathbb{Z}^I, \qquad M \mapsto (\dim M_i)_i$$

and are given by

$$\langle \mathbf{d}, \mathbf{d}'\rangle = {}^t\mathbf{d}\,(Id - C)\,\mathbf{d}', \qquad (\mathbf{d}, \mathbf{d}') = {}^t\mathbf{d}\,A\,\mathbf{d}'.$$

Note that $\dim \mathfrak{M}_{Q,\mathbf{d}} = -\langle \mathbf{d}, \mathbf{d}\rangle$. Now, A is a (symmetric) generalized Cartan matrix in the sense of Kac [1985], to which is attached a Kac-Moody algebra $\mathfrak{g}_Q$. The Euler lattice $(\mathbb{Z}^I, (\ ,\ ))$ of $\mathrm{Rep}_{\mathbf{k}} Q$ is identified to the root lattice[3] $Q_{\mathfrak{g}_Q}$ of $\mathfrak{g}_Q$ together with its standard Cartan pairing, via the map $\mathbb{Z}^I \to Q_{\mathfrak{g}_Q}, \mathbf{d} \mapsto \sum_i d_i\alpha_i$ (here α_i are the simple roots of $\mathfrak{g}_Q$). Accordingly, we denote by $\mathfrak{g}_Q = \bigoplus_{\mathbf{d}} \mathfrak{g}_Q[\mathbf{d}]$ the root space decomposition of $\mathfrak{g}_Q$.

2.2 Kac's theorem and the constant term conjecture. Kac proved that $A_{Q,\mathbf{d}}(t) \neq 0$ if and only if $\mathbf{d}$ belongs to the root system Δ^+ of $\mathfrak{g}_Q$; this generalizes the famous theorem of Gabriel [1972] which concerns the case of finite-dimensional $\mathfrak{g}_Q$. Moreover, he made the following conjecture, which was later proved by Hausel (see Crawley-Boevey and Van den Bergh [2004] for the case of indivisible $\mathbf{d}$):

Theorem 2.1 (Hausel [2010]). *For any* $\mathbf{d} \in \Delta^+$, *we have* $A_{Q,\mathbf{d}}(0) = \dim \mathfrak{g}_Q[\mathbf{d}]$.

[3] we apologize for the unfortunate –yet unavoidable– clash between the traditional notations for the root lattice and for the quiver

We will sketch a proof of Theorems 1.2 and 2.1 using cohomological Hall algebras in Section 4.4. These proofs are different from, but related to the original proofs of Hausel et al.

In the remainder of Section 2, we discuss two (related) constructions of the Kac-Moody algebra $\mathfrak{g}_Q$: as the *spherical Hall algebra* of the categories $\mathrm{Rep}_{\mathbf{k}}(Q)$ for $\mathbf{k}$ a finite field, or in terms of the complex geometry of the *Lusztig nilpotent varieties* $\Lambda_{Q,\mathbf{d}}$.

2.3 Ringel-Hall algebras. It is natural at this point to ask for an actual construction of $\mathfrak{g}_Q$ using the moduli stacks $\mathfrak{M}_{Q,\mathbf{d}}$ of representations of Q. This was achieved by Ringel [1990] and Green [1995] using the *Hall* (or *Ringel-Hall*) algebra of $\mathrm{Rep}_{\mathbf{k}} Q$. Let $\mathbf{k}$ be again a finite field and set $\mathfrak{M}_Q = \bigsqcup_{\mathbf{d}} \mathfrak{M}_{Q,\mathbf{d}}$. The set of $\mathbf{k}$-points $\mathfrak{M}_Q(\mathbf{k}) = \bigsqcup_{\mathbf{d}} \mathfrak{M}_{Q,\mathbf{d}}(\mathbf{k})$ is by construction the set of $\mathbf{k}$-representations of Q (up to isomorphism). Put

$$\mathbf{H}_Q = \bigoplus_{\mathbf{d}} \mathbf{H}_Q[\mathbf{d}], \qquad \mathbf{H}_Q[\mathbf{d}] := \mathrm{Fun}(\mathfrak{M}_{Q,\mathbf{d}}(\mathbf{k}), \mathbb{C})$$

and consider the following convolution diagram

$$\mathfrak{M}_Q \times \mathfrak{M}_Q \xleftarrow{\;q\;} \widetilde{\mathfrak{M}}_Q \xrightarrow{\;p\;} \mathfrak{M}_Q \tag{2-2}$$

where $\widetilde{\mathfrak{M}}_Q$ is the stack parametrizing short exact sequences

$$0 \longrightarrow M \longrightarrow R \longrightarrow N \longrightarrow 0$$

in $\mathrm{Rep}_{\mathbf{k}} Q$ (alternatively, inclusions $M \subset R$ in $\mathrm{Rep}_{\mathbf{k}} Q$); the map p assigns to a short exact sequence as above its middle term R; the map q assigns to it its end terms (N, M). The stack $\widetilde{\mathfrak{M}}_Q$ can be seen as parametrizing extensions between objects in $\mathrm{Rep}_{\mathbf{k}} Q$. The map p is proper and, because $\mathrm{Rep}_{\mathbf{k}} Q$ is of homological dimension one, the map q is a stack vector bundle whose restriction to $\mathfrak{M}_{Q,\mathbf{d}} \times \mathfrak{M}_{Q,\mathbf{d}'}$ is of rank $-\langle \mathbf{d}', \mathbf{d} \rangle$. Following Ringel and Green, we put $v = (\#\mathbf{k})^{\frac{1}{2}}$, let $\mathbf{K} = \mathbb{C}[k_i^{\pm 1}]_i$ be the group algebra of $\mathbb{Z}^I$ and we equip $\widetilde{\mathbf{H}}_Q := \mathbf{H}_Q \otimes \mathbf{K}$ with the structure of a $\mathbb{N}^I$-graded bialgebra by setting

$$f \cdot g = v^{\langle \mathbf{d}, \mathbf{d}' \rangle} p_! q^*(f \otimes g), \qquad k_{\mathbf{n}} f k_{\mathbf{n}}^{-1} = v^{(\mathbf{d}, \mathbf{n})} f \tag{2-3}$$

$$\Delta(f) = \sum_{\mathbf{d}' + \mathbf{d}'' = \mathbf{d}} v^{\langle \mathbf{d}', \mathbf{d}'' \rangle} (q_{\mathbf{d}', \mathbf{d}''})_! p^*_{\mathbf{d}', \mathbf{d}''}(f) \qquad \Delta(k_i) = k_i \otimes k_i \tag{2-4}$$

for $f \in \mathbf{H}_Q[\mathbf{d}]$, $g \in \mathbf{H}_Q[\mathbf{d}']$, $\mathbf{n} \in \mathbb{Z}^I$ and $i \in I$. The bialgebra $\widetilde{\mathbf{H}}_Q$ is equipped with the nondegenerate Hopf pairing

$$(f k_{\mathbf{d}}, g k_{\mathbf{d}'}) = v^{(\mathbf{d}, \mathbf{d}')} \int_{\mathfrak{M}_Q} f \overline{g}, \qquad \forall\ f, g \in \mathbf{H}_Q \tag{2-5}$$

(see Schiffmann [2012] for more details and references). Denoting by $\{\epsilon_i\}_i$ the canonical basis of $\mathbb{Z}^I$ we have $\mathfrak{M}_{\epsilon_i} \simeq B\mathbb{G}_m$ for all i since $\mathfrak{M}_{\epsilon_i}$ has as unique object the simple object S_i of dimension ϵ_i. We define the *spherical* Hall algebra $\widetilde{\mathbf{H}}_Q^{\mathrm{sph}}$ as the subalgebra of $\widetilde{\mathbf{H}}_Q$ generated by $\mathbf{K}$ and the elements 1_{S_i} for $i \in I$. It is a sub-bialgebra.

Theorem 2.2 (Ringel [1990] and Green [1995]). *The assignement $E_i \mapsto 1_{S_i}, K_i^{\pm 1} \mapsto k_i^{\pm 1}$ for $i \in I$ induces an isomorphism $\Psi : U_v(\mathfrak{b}_Q) \xrightarrow{\sim} \widetilde{\mathbf{H}}_Q^{sph}$ between the positive Borel subalgebra of the Drinfeld-Jimbo quantum enveloping algebra $U_v(\mathfrak{g}_Q)$ and the spherical Hall algebra of Q.*

One may recover the entire quantum group $U_v(\mathfrak{g}_Q)$ as the (reduced) Drinfeld double $\mathbf{D}\widetilde{\mathbf{H}}_Q^{\mathrm{sph}}$ of $\widetilde{\mathbf{H}}_Q$ (see e.g. Schiffmann [2012, Lec. 5]).

2.4 Lusztig nilpotent variety. Let $T^*\mathfrak{M}_Q = \bigsqcup_{\mathbf{d}} T^*\mathfrak{M}_{Q,\mathbf{d}}$ be the cotangent[4] stack of $\mathfrak{M}_Q$. This may be realized explicitly as follows. Let $\overline{Q} = (I, \Omega \sqcup \Omega^*)$ be the double of Q, obtained by replacing every arrow h in Q by a pair (h, h^*) of arrows going in opposite directions. The preprojective algebra Π_Q is the quotient of the path algebra $\mathbf{k}\overline{Q}$ by the two-sided ideal generated by $\sum_{h \in \Omega}[h, h^*]$. Unless Q is of finite type, the abelian category $\mathrm{Rep}_{\mathbf{k}}\Pi_Q$ is of global dimension two. The stack of $\mathbf{k}$-representations of Π_Q is naturally identified with $T^*\mathfrak{M}_Q$. We say that a representation M of Π_Q is *nilpotent* if there exists a filtration $M \supset M_1 \supset \cdots \supset M_l = \{0\}$ for which $\Pi_Q^+(M_i) \subseteq M_{i+1}$, where $\Pi_Q^+ \subset \Pi_Q$ is the augmentation ideal. Following Lusztig, we define the *nilpotent variety* (or stack) $\Lambda_Q = \bigsqcup_{\mathbf{d}} \Lambda_{Q,\mathbf{d}} \subset T^*\mathfrak{M}_Q$ as the substack of nilpotent representations of Π_Q.

Theorem 2.3 (Kashiwara and Saito [1997]). *The stack Λ_Q is a lagrangian substack of $T^*\mathfrak{M}_Q$ and for any $\mathbf{d} \in \mathbb{N}^I$ we have $\#Irr(\Lambda_{Q,\mathbf{d}}) = dim\, U(\mathfrak{n}_Q)[\mathbf{d}]$.*

The above theorem is strongly motivated by Lusztig's geometric lift of the spherical Hall algebra to a category of perverse sheaves on $\mathfrak{M}_Q$, yielding the *canonical basis* $\mathbf{B}$ of $U_v(\mathfrak{n}_Q)$, see Lusztig [1991]. It turns out that Λ_Q is precisely the union of the singular support of Lusztig's simple perverse sheaves, hence the above may be seen as a microlocal avatar of Lusztig's construction. Kashiwara and Saito prove much more, namely they equip $Irr(\Lambda)$ with the combinatorial structure of a *Kashiwara crystal* which they identify as the crystal $B(\infty)$ of $U(\mathfrak{n}_Q)$. This also yields a canonical bijection between $\mathbf{B}$ and $Irr(\Lambda_Q)$.

2.5 Arbitrary quivers and nilpotent Kac polynomials. What happens when the quiver Q *does* have edge loops, such as the g-loop quiver S_g ? This means that the matrix of

[4] to be precise, we only consider the underived, or H^0-truncation, of the cotangent stacks here

the Euler form may have some (even) nonpositive entries on the diagonal, hence it is associated to a *Borcherds* algebra rather than to a Kac-Moody algebra. Accordingly, we call *real*, resp. *isotropic*, resp. *hyperbolic* a vertex carrying zero, resp. one, resp. at least two edge loops, and in the last two cases we say that the vertex is *imaginary*. One may try to get Hall-theoretic constructions of the Borcherds algebra $\mathfrak{g}_Q$ associated to Q (see e.g. Kang and Schiffmann [2006] and Kang, Kashiwara, and Schiffmann [2009]) but the best thing to do in order to get a picture as close as possible to the one in **2.1–2.4.** seems to be to consider instead a slightly larger algebra $\mathfrak{g}_Q^B$ defined by Bozec which has as building blocks the usual $\mathfrak{sl}_2$ for real roots, the Heisenberg algebra H for isotropic roots and a free Lie algebra H' with one generator in each degree for hyperbolic roots, see Bozec [2015].

The Hall algebra $\widetilde{\mathbf{H}}_Q$ is defined just as before, but we now let $\widetilde{\mathbf{H}}_Q^{\mathrm{sph}}$ be the subalgebra generated by elements 1_{S_i} for i real vertices and by the characteristic functions $1_{\mathfrak{M}_{l\epsilon_i}}$ for i imaginary and $l \in \mathbb{N}$. The stack of nilpotent representations of Π_Q is not lagrangian anymore in general, we consider instead the stack of *seminilpotent* representations, i.e. representations M for which there exists a filtration $M \supset M_1 \supset \cdots \supset M_l = \{0\}$ such that $h(M_i/M_{i+1}) = 0$ for $h \in \Omega$ or for $h \in \Omega^*$ *not* an edge loop.

Theorem 2.4 (Bozec [2015, 2016] and Kang [2018]). *The following hold:*

i) $\widetilde{\mathbf{H}}_Q^{sph}$ *is isomorphic to* $U_v(\mathfrak{b}_Q^B)$,

ii) The stack Λ_Q *of seminilpotent representations is lagrangian in* $T^*\mathfrak{M}_Q$. *For any* $\mathbf{d} \in \mathbb{N}^I$ *we have* $\#Irr(\Lambda_{Q,\mathbf{d}}) = dim\, U(\mathfrak{n}_Q^B)[\mathbf{d}]$.

The theories of canonical bases and crystal graphs also have a natural extension to this setting. What about the relation to Kac polynomials ? Here as well we need some slight variation; let us call 1-*nilpotent* a representation M of Q in which the edge loops at any given imaginary vertex i generate a nilpotent (associative) subalgebra of $\mathrm{End}(M_i)$. Of course any representation of a quiver with no edge loops is 1-nilpotent.

Theorem 2.5 (Bozec, Schiffmann, and Vasserot [2017]). *For any* $\mathbf{d}$ *there exists a (unique) polynomial* $A^{nil}_{Q,\mathbf{d}} \in \mathbb{Z}[t]$, *independent of the orientation of* Q, *such that for any finite field* $\mathbf{k}$ *the number of isomorphism classes of absolutely indecomposable* 1*-nilpotent* $\mathbf{k}$*-representations of* Q *of dimension* $\mathbf{d}$ *is equal to* $A^{nil}_{Q,\mathbf{d}}(\#\mathbf{k})$. *Moreover, for any* $\mathbf{d} \in \mathbb{N}^I$ *we have*

i) $A^{nil}_{Q,\mathbf{d}}(1) = A_{Q,\mathbf{d}}(1)$,

ii) $A^{nil}_{Q,\mathbf{d}}(0) = dim\, \mathfrak{g}_Q^B[\mathbf{d}]$.

We will sketch a proof of the fact that $A^{\mathrm{nil}}_{Q,\mathbf{d}}(t) \in \mathbb{N}[t]$ in Section 4.4.

3 Quivers, Kac polynomials and graded Borcherds algebras I –counting cuspidals–

Starting form a Kac-Moody algebra $\mathfrak{g}$ (or its variant defined by Bozec) we build a quiver Q by orienting the Dynkin diagram of $\mathfrak{g}$ in any fashion, and obtain a realization of $U_v(\mathfrak{b})$ as the spherical Hall algebra $\widetilde{\mathbf{H}}_Q^{\mathrm{sph}}$; now, we may ask the following question: what is the relation between $\mathfrak{g}$ and the *full* Hall algebra ? Here's a variant of this question: Kac's constant term conjecture gives an interpretation of the constant term of the Kac polynomial $A_{Q,\mathbf{d}}(t)$; is there a similar Lie-theoretic meaning for the *other* coefficients ? As we will explain in **3.3.** below, the two questions turn out to have a beautiful common conjectural answer: the full Hall algebra $\widetilde{\mathbf{H}}_Q$ is related to a *graded* Borcherds algebra $\widetilde{\mathfrak{g}}_Q$ whose *graded* multiplicities $\dim_{\mathbb{Z}} \widetilde{\mathfrak{g}}_Q[\mathbf{d}] = \sum_l \dim \widetilde{\mathfrak{g}}_Q[\mathbf{d}, l]t^l$ are equal to $A_{Q,\mathbf{d}}(t)$. What's more, $\widetilde{\mathfrak{g}}_Q$, like $\mathfrak{g}_Q$, is independent of the choice of the orientation of the quiver and is hence *canonically* attached to $\mathfrak{g}_Q$. Although the above makes sense for an arbitrary quiver Q, there is an entirely similar story in the nilpotent setting (better suited for a quiver with edge loops), replacing $A_{Q,\mathbf{d}}(t)$ by $A^{\mathrm{nil}}_{Q,\mathbf{d}}(t)$ and $\mathbf{H}_Q$ by the subalgebra $\mathbf{H}_Q^{\mathrm{nil}}$ of functions on the stack $\mathfrak{M}_Q^{\mathrm{nil}}$ of 1-nilpotent representations of Q. This is then conjecturally related to a graded Borcherds algebra $\widetilde{\mathfrak{g}}_Q^{\mathrm{nil}}$.

3.1 The full Hall algebra.

The first general result concerning the structure of the full Hall algebra $\widetilde{\mathbf{H}}_Q$ is due to Sevenhant and Van den Bergh. Let us call an element $f \in \mathbf{H}_Q$ is *cuspidal* if $\Delta(f) = f \otimes 1 + k_f \otimes 1$. Denote by $\mathbf{H}_Q^{\mathrm{cusp}} = \bigoplus_{\mathbf{d}} \mathbf{H}_Q^{\mathrm{cusp}}[\mathbf{d}]$ the space of cuspidals, and set $C_{Q/\mathbf{k},\mathbf{d}} = \dim\ \mathbf{H}_Q^{\mathrm{cusp}}[\mathbf{d}]$. We write $Q/\mathbf{k}$ instead of Q to emphasize the dependance on the field. Consider the infinite Borcherds–Cartan data $(A_{Q/\mathbf{k}}, \mathbf{m}_{Q/\mathbf{k}})$, with $\mathbb{N}^I \times \mathbb{N}^I$ Cartan matrix $A_{Q/\mathbf{k}} = (a_{\mathbf{d},\mathbf{d}'})_{\mathbf{d},\mathbf{d}'}$ and charge function $\mathbf{m}_{Q/\mathbf{k}} : \mathbb{N}^I \to \mathbb{N}$ defined as follows:

$$a_{\mathbf{d},\mathbf{d}'} = (\mathbf{d}, \mathbf{d}'), \qquad \mathbf{m}_{Q/\mathbf{k}}(\mathbf{d}) = C_{Q/\mathbf{k},\mathbf{d}} \qquad \forall\ \mathbf{d}, \mathbf{d}' \in \mathbb{N}^I.$$

Denote by $\widetilde{\mathfrak{g}}_{Q/\mathbf{k}}$ the Borcherds algebra associated to $(A_{Q/\mathbf{k}}, \mathbf{m}_{Q/\mathbf{k}})$.

Theorem 3.1 (Sevenhant and Van Den Bergh [2001]). *The Hall algebra $\mathbf{H}_{Q/\mathbf{k}}$ is isomorphic to the positive nilpotent subalgebra $U_v(\widetilde{\mathfrak{n}}_{Q/\mathbf{k}})$ of the Drinfeld-Jimbo quantum enveloping algebra of $\widetilde{\mathfrak{g}}_{Q/\mathbf{k}}$.*

In other words, the cuspidal element span the spaces of *simple root vectors* for the Hall algebra $\mathbf{H}_{Q/\mathbf{k}}$. Note that adding the 'Cartan' $\mathbf{K}$ to $\mathbf{H}_Q$ will only produce a quotient of $U_v(\widetilde{\mathfrak{b}}_{Q/\mathbf{k}})$, whose own Cartan subalgebra is not finitely generated.

3.2 Counting cuspidals for quivers. To understand the structure of $\widetilde{\mathfrak{g}}_{Q/\mathbf{k}}$ better, we first need to understand the numbers $C_{Q/\mathbf{k},\mathbf{d}}$. In this direction, we have the following theorem, proved by Deng and Xiao for quivers with no edge loops, and then extended to the general case in Bozec and Schiffmann [2017]:

Theorem 3.2 (Deng and Xiao [2003] and Bozec and Schiffmann [2017])**.** *For any $Q = (I,\Omega)$ and $\mathbf{d} \in \mathbb{N}^I$ there exists a (unique) polynomial $C_{Q,\mathbf{d}}(t) \in \mathbb{Q}[t]$ such that for any finite field* $\mathbf{k}$*, $C_{Q/\mathbf{k},\mathbf{d}} = C_{Q,\mathbf{d}}(\#\mathbf{k})$.*

Let us say a word about the proof. By the Krull-Schmidt theorem, there is an isomorphism of $\mathbb{N}^I$-graded vector spaces between $\mathbf{H}_{Q/\mathbf{k}}$ and $Sym(\bigoplus_{\mathbf{d}} Fun(\mathfrak{M}^{ind}_{Q/\mathbf{k},\mathbf{d}}(\mathbf{k}),\mathbb{C}))$, where $\mathfrak{M}^{ind}_{Q/\mathbf{k},\mathbf{d}}(\mathbf{k})$ is the groupoid of indecomposable $\mathbf{k}$-representations of Q of dimension $\mathbf{d}$. This translates into the equality of generating series

$$\sum_{\mathbf{d}}(\dim \mathbf{H}_{Q/\mathbf{k}}[\mathbf{d}])z^d = \operatorname{Exp}_z\left(\sum_{\mathbf{d}} I_{Q,\mathbf{d}}(t)z^d\right)_{|t=\#\mathbf{k}} \tag{3-1}$$

where Exp_z stands for the plethystic exponential with respect to the variable z, see e.g. Mozgovoy [2011]. From the PBW theorem,

$$\sum_{\mathbf{d}}(\dim U(\widetilde{\mathfrak{n}}_{Q/\mathbf{k}})[\mathbf{d}])z^d = \operatorname{Exp}_z\left(\sum_{\mathbf{d}}(\dim \widetilde{\mathfrak{n}}_{Q/\mathbf{k}}[\mathbf{d}])z^d\right) \tag{3-2}$$

from which it follows that $\dim \widetilde{\mathfrak{n}}_{Q/\mathbf{k}}[\mathbf{d}] = I_{Q,\mathbf{d}}(\#\mathbf{k})$ for any $\mathbf{d}$. We are in the following situation: we know the character[5] of the Borcherds algebra $\widetilde{\mathfrak{g}}_{Q/\mathbf{k}}$ and from that we want to determine its Borcherds-Cartan data. This amounts to inverting the Borcherds character (or denominator) formula, which may be achieved in each degree by some finite iterative process[6]. Following through this iterative process and using the fact that $I_{Q,d}(\#\mathbf{k})$ is a polynomial in $\#\mathbf{k}$, one checks that the Borcherds-Cartan data, in particular the charge function $\mathbf{m}_{Q/\mathbf{k}}(\mathbf{d}) = C_{Q/\mathbf{k},\mathbf{d}}$ is also polynomial in $\#\mathbf{k}$. This process is constructive, but as far as we know no closed formula is known in general (see Bozec and Schiffmann [2017] for some special cases, such as totally negative quivers like S_g, $g > 1$). And in particular, the structure (and the size) of $\mathbf{H}_{Q/\mathbf{k}}$ depends heavily on $\mathbf{k}$.

[5] there is a subtle point here: $\widetilde{\mathfrak{g}}_{Q/\mathbf{k}}$ is $\mathbb{Z}^{(\mathbb{N}^I)}$-graded, but we only know the character after projection of the grading to $\mathbb{N}^I$.

[6] here we use the fact that the Weyl group of $\widetilde{\mathfrak{g}}_{Q/\mathbf{k}}$ only depends on the subset of real simple roots, which is the same as that of $\mathfrak{g}_Q$.

3.3 Counting absolutely cuspidals for quivers. In order to get a more canonical structure out of $\mathbf{H}_{Q/\mathbf{k}}$ one is tempted to view the polynomial $C_{Q,\mathbf{d}}(t)$ as giving the graded dimension of some vector space. Unfortunately, as can readily be seen for the Jordan quiver S_1 and $\mathbf{d} = 2$, the polynomial $C_{Q,\mathbf{d}}(t)$ fails to be integral or positive in general. This is familiar: the same things happens for Kac's A and I polynomials and one might guess that the better thing to count would be *absolutely* cuspidal elements of $\mathbf{H}_{Q/\mathbf{k}}$. Unfortunately (and contrary to the case of curves), there is at the moment no known definition of an absolutely cuspidal element of $\mathbf{H}_{Q/\mathbf{k}}$! One way around this is to use a well-known identity between Kac A-polynomials and I-polynomials to rewrite (3-1) as

(3-3)
$$\sum_{\mathbf{d}} \dim(U(\widetilde{\mathfrak{n}}_{Q/\mathbf{k}})[\mathbf{d}])z^d = \sum_{\mathbf{d}} \dim(H_{Q/\mathbf{k}}[\mathbf{d}])z^d = \operatorname{Exp}_{z,t}\left(\sum_{\mathbf{d}} A_{Q,\mathbf{d}}(t)z^d\right)_{|t=\#\mathbf{k}}$$

and interpret the right hand side as the character of the enveloping algebra of (the nilpotent subalgebra of) a putative *graded* Borcherds algebra $\widetilde{\mathfrak{g}}_Q$ whose graded character is given by the polynomials $A_{Q,\mathbf{d}}(t)$

$$\dim_{\mathbb{Z}} \widetilde{\mathfrak{g}}_Q[\mathbf{d}] = A_{Q,\mathbf{d}}(t). \tag{3-4}$$

Taking this as a working hypothesis, we can run through the above iterative process (in the graded sense) to get a well-defined family of polynomials $C^{\text{abs}}_{Q,\mathbf{d}}(t)$.

Theorem 3.3 (Bozec and Schiffmann [2017]). *For any Q and $\mathbf{d}$ we have $C^{abs}_{Q,\mathbf{d}}(t) \in \mathbb{Z}[t]$. In addition, we have $C^{abs}_{Q,\mathbf{d}}(t) = C_{Q,\mathbf{d}}(t)$ for any hyperbolic $\mathbf{d}$ (i.e. $(\mathbf{d},\mathbf{d}) < 0$) while for any indivisible isotropic $\mathbf{d}$,*

$$\operatorname{Exp}_z\left(\sum_{l\geq 1} C_{Q,l\mathbf{d}}(t)z^l\right) = \operatorname{Exp}_{t,z}\left(\sum_{l\geq 1} C^{abs}_{Q,l\mathbf{d}}(t)z^l\right)$$

Our definition of $C^{\text{abs}}_{Q,\mathbf{d}}(t)$ was motivated by the putative existence of a graded Borcherds algebra $\widetilde{\mathfrak{g}}_Q$. This existence is *a posteriori* equivalent to the following conjecture:

Conjecture 3.4. *For any Q and $\mathbf{d}$ we have $C^{abs}_{Q,\mathbf{d}} \in \mathbb{N}[t]$.*

We stress that the above positivity conjecture is strictly stronger than the positivity of Kac's A-polynomials. Indeed by construction Kac polynomials are integral positive polynomials in the $C^{\text{abs}}_{Q,\mathbf{d}}(t)$. We will give in Sections 4.4 and 4.6 two other (conjectural) constructions of the Lie algebra $\widetilde{\mathfrak{g}}_Q$, based on the geometry of $T^*\mathfrak{M}_Q, \Lambda_Q$ and on Nakajima quiver varieties respectively, either of which would imply the above positivity conjecture. We finish this section with another (somewhat imprecise) conjecture, which is a quiver analog of a conjecture of Kontsevich and Deligne, see Section 6.1.

Conjecture 3.5. *For any* Q, $\mathbf{d}$ *there exists a 'natural' algebraic variety* $\mathcal{C}_{Q,\mathbf{d}}$ *defined over* $\mathbb{Z}$ *such that for any finite field* $\mathbf{k}$ *we have* $C^{abs}_{Q,\mathbf{d}}(\#\mathbf{k}) = \#\mathcal{C}_{Q,\mathbf{d}}(\mathbf{k})$.

There are obvious variants of this conjecture, replacing $\mathcal{C}_{Q,\mathbf{d}}$ by a complex algebraic variety and the point count by the Poincaré polynomial, fixing the characteristic of $\mathbf{k}$, etc. *Examples:* For Q an affine quiver, we have $C^{\text{abs}}_{Q,\mathbf{d}}(t) = 1, t$ or 0 according to whether $\mathbf{d} \in \{\epsilon_i\}_i$, $\mathbf{d} \in \mathbb{N}\delta$ for δ the indivisible imaginary root, or neither of the above. For $Q = S_3$ and $\mathbf{d} = 3$ we have

$$C^{\text{abs}}_{S_g,3} = \frac{t^{9g-3} - t^{5g+2} - t^{5g-2} - t^{5g-3} + t^{3g+2} + t^{3g-2}}{(t^2-1)(t^3-1)}.$$

3.4 More questions than answers. As we have seen, there is a method for counting cuspidal (or even 'absolutely' cuspidal) functions, via Kac polynomials and the Borcherds character formula. But is it possible to get a geometric parametrization of these cuspidals (either in the sense of Conjecture 3.5, or as in the Langlands program in terms of some 'spectral data' [7]) ? Is it possible to explicitly construct cuspidal functions ? Can one lift a suitable basis of the space of cuspidal functions to some perverse sheaves on $\mathfrak{M}_Q$? In other words, is there a theory of canonical bases for $U_v(\widetilde{\mathbf{n}}_Q)$? If so, can one describe explicitly the Ext-algebra of that category of perverse sheaves ? Equivalently, is there a analog, for $\widetilde{\mathfrak{g}}_Q$, of the Khovanov-Lauda-Rouquier algebra, see Khovanov and Lauda [2009] and Rouquier [2012] ?

4 Quivers, Kac polynomials and graded Borcherds algebras II –Cohomological Hall algebras and Yangians–

The previous section offered a conjectural definition of Lie algebras $\widetilde{\mathfrak{g}}_Q, \widetilde{\mathfrak{g}}^{\text{nil}}_Q$ whose graded multiplicities are $A_{Q,\mathbf{d}}(t)$, $A^{\text{nil}}_{Q,\mathbf{d}}(t)$, but the construction –involving something like a 'generic form' for the full Hall algebras $\mathbf{H}_Q, \mathbf{H}^{\text{nil}}_Q$– was somewhat roundabout. In this section we describe a geometric construction, in terms of the cohomology of $T^*\mathfrak{M}_Q$ or Λ_Q, of algebras –the *cohomological Hall algebras*– which are deformations of the enveloping algebras $U(\bar{\mathfrak{n}}_Q[u]), U(\bar{\mathfrak{n}}^{\text{nil}}_Q[u])$ for certain graded Lie algebras $\bar{\mathfrak{n}}_Q, \bar{\mathfrak{n}}^{\text{nil}}_Q$ satisfying $\dim_{\mathbb{Z}} \bar{\mathfrak{n}}_Q = A_{Q,\mathbf{d}}(t), \dim_{\mathbb{Z}} \bar{\mathfrak{n}}^{\text{nil}}_Q = A^{\text{nil}}_{Q,\mathbf{d}}(t)$. Of course, it is expected that $\bar{\mathfrak{n}}_Q, \bar{\mathfrak{n}}^{\text{nil}}_Q$ are positive halves of Borcherds algebras $\bar{\mathfrak{g}}_Q, \bar{\mathfrak{g}}^{\text{nil}}_Q$ and thus coincide with $\widetilde{\mathfrak{g}}_Q, \widetilde{\mathfrak{g}}^{\text{nil}}_Q$. One nice output of this construction is that it yields for free a whole family of representations (in the cohomology of Nakajima quiver varieties). In this section, $\mathbf{k} = \mathbb{C}$.

[7] a preliminary question: what plays the role of Hecke operators in this context?

4.1 Geometry of the nilpotent variety. Let us begin by mentioning a few remarkable geometric properties of the (generally very singular) stacks $T^*\mathfrak{M}_Q, \Lambda_Q$ introduced in **2.4.** and **2.5.**. Here $Q = (I, \Omega)$ is an arbitrary quiver. Denote by $T = (\mathbb{C}^*)^2$ the two-dimensional torus acting on Λ_Q and $T^*\mathfrak{M}_Q$ by scaling the arrows in Ω and Ω^* respectively.

Theorem 4.1 (Schiffmann and Vasserot [2017a] and Davison [2016])**.** *The T-equivariant Borel-Moore homology space $H_*^T(\Lambda_Q, \mathbb{Q}) = \bigoplus_{\mathbf{d}} H_*^T(\Lambda_{Q,\mathbf{d}}, \mathbb{Q})$ is even, pure, and free over $H_T^*(pt, \mathbb{Q})$. The same holds for $T^*\mathfrak{M}_Q$.*

We will only consider (co)homology with $\mathbb{Q}$-coefficients and drop the $\mathbb{Q}$ in the notation. The above theorem for Λ_Q is proved by constructing a suitable compactification of $\Lambda_{Q,\mathbf{d}}$ for each $\mathbf{d}$, itself defined in terms of Nakajima varieties, see Schiffmann and Vasserot [2017a] and Nakajima [1994]. The second part of the theorem relies on dimensional reduction from a 3d Calabi-Yau category, see Davison [2016]. Kac polynomials are intimately related to the cohomology of Λ_Q and $T^*\mathfrak{M}_Q$. More precisely,

Theorem 4.2 (Bozec-S.-Vasserot, Davison, Mozgovoy)**.** *The Poincaré polynomials of Λ_Q and $T^*\mathfrak{M}_Q$ are respectively given by:*

$$\sum_{i,\mathbf{d}} dim\ H_{2i}(\Lambda_{Q,\mathbf{d}}) t^{\langle \mathbf{d},\mathbf{d}\rangle + i} z^{\mathbf{d}} = Exp\left(\sum_{\mathbf{d}} \frac{A^{nil}_{Q,\mathbf{d}}(t^{-1})}{1-t^{-1}} z^{\mathbf{d}}\right) \tag{4-1}$$

$$\sum_{i,\mathbf{d}} dim\ H_{2i}(T^*\mathfrak{M}_{Q,\mathbf{d}}) t^{\langle \mathbf{d},\mathbf{d}\rangle + i} z^{\mathbf{d}} = Exp\left(\sum_{\mathbf{d}} \frac{A_{Q,\mathbf{d}}(t)}{1-t^{-1}} z^{\mathbf{d}}\right). \tag{4-2}$$

Equality (4-1) is proved in Bozec, Schiffmann, and Vasserot [2017] and hinges again on some partial compactification of the stacks $\Lambda_{Q,\mathbf{d}}$ defined in terms of Nakajima quiver varieties. The second equality (4-2) is obtained by combining the purity results of Davison [2016] with the point count of Mozgovoy [2011]. In some sense, (4-1) and (4-2) are Poincaré dual to each other. Note that $\dim \Lambda_{Q,\mathbf{d}} = -\langle \mathbf{d}, \mathbf{d}\rangle$ hence (4-1) is a series in $\mathbb{C}[[t^{-1}]][z]$. Taking the constant term in t^{-1} yields $\sum_{\mathbf{d}} \#Irr(\Lambda_{\mathbf{d}}) z^{\mathbf{d}} = Exp(\sum_{\mathbf{d}} A^{\text{nil}}_{\mathbf{d}}(0) z^{\mathbf{d}})$, in accordance with Theorem 2.4 ii), Theorem 2.5 ii). In other terms, (4-1) provides a geometric interpretation (as the Poincaré polynomial of some stack) of the *full* Kac polynomial $A^{\text{nil}}_{Q,\mathbf{d}}(t)$ rather than just its constant term and passing from $A^{\text{nil}}_{Q,\mathbf{d}}(0)$ to $A^{\text{nil}}_{Q,\mathbf{d}}(t)$ essentially amounts to passing from $H_{top}(\Lambda_{Q,\mathbf{d}})$ to $H_*(\Lambda_{Q,\mathbf{d}})$. Of course, (4-2) has a similar interpretation.

4.2 Cohomological Hall algebras. The main construction of this section is the following:

Theorem 4.3 (Schiffmann and Vasserot [2013a] and Schiffmann and Vasserot [2017a]). *The spaces $H_*^T(T^*\mathfrak{M}_Q)$ and $H_*^T(\Lambda_Q)$ carry natural $\mathbb{Z}\times\mathbb{N}^I$-graded algebra structures. The direct image morphism $i_* : H_*^T(\Lambda_Q) \to H_*^T(T^*\mathfrak{M}_Q)$ is an algebra homomorphism.*

Let us say a few words about how the multiplication map

$$H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}_1}) \otimes H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}_2}) \to H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}_1+\mathbf{d}_2})$$

is defined. There is a convolution diagram

$$T^*\mathfrak{M}_{Q,\mathbf{d}_1} \times T^*\mathfrak{M}_{Q,\mathbf{d}_2} \xleftarrow{\ q\ } Z_{\mathbf{d}_1,\mathbf{d}_2} \xrightarrow{\ p\ } T^*\mathfrak{M}_{Q,\mathbf{d}_1+\mathbf{d}_2}$$

similar to (2-2), where $Z_{\mathbf{d}_1,\mathbf{d}_2}$ is the stack of inclusions $\bar{M} \subset \bar{R}$ with $\bar{M}, \bar{R}$ representations of Π_Q of respective dimensions $\mathbf{d}_2, \mathbf{d}_1+\mathbf{d}_2$, and p and q are the same as in (2-2). The map p is still proper, so that $p_* : H_*^T(Z_{\mathbf{d}_1,\mathbf{d}_2}) \to H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}_1+\mathbf{d}_2})$ is well-defined, but q is not regular anymore and we cannot directly define a Gysin map q^*. Instead, we embedd $Z_{\mathbf{d}_1,\mathbf{d}_2}$ and $T^*\mathfrak{M}_{Q,\mathbf{d}_1} \times T^*\mathfrak{M}_{Q,\mathbf{d}_2}$ into suitable *smooth* moduli stacks of representations of the path algebra of the double quiver $\overline{Q}$ (without preprojective relations) and define a refined Gysin map $q^! : H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}_1} \times T^*\mathfrak{M}_{Q,\mathbf{d}_2}) \to H_*^T(Z_{\mathbf{d}_1,\mathbf{d}_2})$. The multiplication map is then $m = p_*q^!$. Note that it is of cohomological degree $-\langle \mathbf{d}_1,\mathbf{d}_2\rangle - \langle \mathbf{d}_2,\mathbf{d}_1\rangle$; to remedy this, we may (and will) shift the degree of $H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}})$ by $\langle \mathbf{d},\mathbf{d}\rangle$. There is another notion of cohomological Hall algebra due to Kontsevich and Soibelman, associated to any Calabi-Yau category of dimension *three* (Kontsevich and Y. Soibelman [2011]); as shown by Davison, see the appendix to Ren and Y. Soibelman [2017] (see also Yang and Zhao [2016]) using a dimensional reduction argument, the algebras $H_*^T(T^*\mathfrak{M}_Q)$, $H_*^T(\Lambda_Q)$ arise in that context as well. In addition, the above construction of the cohomological Hall algebras can be transposed to any oriented cohomology theory which has proper pushforwards and refined Gysin maps, such as Chow groups, K-theory, elliptic cohomology or Morava K-theory (see Yang and Zhao [2014]). This yields in a uniform way numerous types of quantum algebras with quite different flavors.

Not very much is known about the precise structure of $H_*^T(T^*\mathfrak{M}_Q)$ or $H_*^T(\Lambda_Q)$ in general (and even less so for other types of cohomology theories). However, it is possible to give a simple set of generators; for any $\mathbf{d}$ there is an embedding of $\mathfrak{M}_{Q,\mathbf{d}}$ in $\Lambda_{Q,\mathbf{d}}$ as the zero-section, the image being one irreducible component of $\Lambda_{Q,\mathbf{d}}$.

Theorem 4.4 (Schiffmann and Vasserot [2017a]). *The algebra $H_*^T(\Lambda_Q)$ is generated over $H_T^*(pt)$ by the collection of subspaces*

$$\{H_*^T(\mathfrak{M}_{Q,\epsilon_i}) \mid i \in I^{re} \cup I^{iso}\} \cup \{H_*^T(\mathfrak{M}_{Q,l\epsilon_i}) \mid i \in I^{hyp}, l \in \mathbb{N}\}$$

Note that $\mathfrak{M}_{Q,\mathbf{d}} = E_{\mathbf{d}}/G_{\mathbf{d}}$, where $E_{\mathbf{d}}$ is a certain representation of $G_{\mathbf{d}} = \prod_i GL(\mathbf{d}_i)$, hence $H_*^T(\mathfrak{M}_{Q,\mathbf{d}}) = H^*_{G_{\mathbf{d}}\times T}(pt) \cdot [\mathfrak{M}_{Q,\mathbf{d}}] \simeq \mathbb{Q}[q_1, q_2, c_l(M_i) \mid i \in I, l \leq \mathbf{d}_i]$, where

q_1, q_2 are the equivariant parameters corresponding to T. The above result is for the nilpotent stack $\Lambda_{Q,\mathbf{d}}$, but one can show that $H_*^T(\Lambda_Q) \otimes \mathbb{Q}(q_1, q_2) = H_*^T(T^*\mathfrak{M}_Q) \otimes \mathbb{Q}(q_1, q_2)$ [8] so that the same generation result holds for $H_*^T(T^*\mathfrak{M}_Q) \otimes \mathbb{Q}(q_1, q_2)$.

4.3 Shuffle algebras. One can also give an algebraic model for a certain *localized* form of $H_*^T(T^*\mathfrak{M}_Q)$ (or $H_*^T(\Lambda_Q)$). Namely, let $\overline{E}_{\mathbf{d}} = T^*E_{\mathbf{d}}$ be the vector space of representations of $\overline{Q}$ in $\bigoplus_i \mathbf{k}^{\mathbf{d}_i}$. The direct image morphism $i_* : H_*^T(T^*\mathfrak{M}_{Q,\mathbf{d}}) \to H_*^T(\overline{E}_{\mathbf{d}}/G_{\mathbf{d}})$ is an isomorphism after tensoring by $Frac(H^*_{T\times G_{\mathbf{d}}}(pt))$. What's more, it is possible to equip

$$Sh_Q^{H_*^T} := \bigoplus_{\mathbf{d}} H_*^T(\overline{E}_{\mathbf{d}}/G_{\mathbf{d}}) \simeq \bigoplus_{\mathbf{d}} H^*_{T\times G_{\mathbf{d}}}(pt)$$

with the structure of an associative algebra (such that i_* becomes an algebra morphism), described explicitly as a *shuffle algebra*. More precisely, let us identify

$$H^*_{T\times G_{\mathbf{d}}}(pt) \simeq \mathbb{Q}[q_1, q_2, z_{i,l} \mid i \in I, l \leq \mathbf{d}_i]^{W_{\mathbf{d}}} \tag{4-3}$$

where $W_{\mathbf{d}} = \prod_i \mathfrak{S}_{\mathbf{d}_i}$ is the Weyl group of $G_{\mathbf{d}}$. To unburden the notation, we will collectively denote the variables $z_{i,1}, \ldots, z_{i,\mathbf{d}_i}$ (for all $i \in I$) by $\underline{z}_{[1,\mathbf{d}]}$, and use obvious variants of that notation. Also, we will regard q_1, q_2 as scalars and omit them from the notation. Fix dimension vectors $\mathbf{d}, \mathbf{e}$ and put $\mathbf{n} = \mathbf{d}+\mathbf{e}$. For two integers r, s we denote by $Sh_{r,s} \subset \mathfrak{S}_{r+s}$ the set of (r,s)-shuffles, i.e. permutations σ satisfying $\sigma(i) < \sigma(j)$ for $1 \leq i < j \leq r$ and $r < i < j \leq r+s$, and we put $Sh_{\mathbf{d},\mathbf{e}} = \prod_i Sh_{\mathbf{d}_i,\mathbf{e}_i} \subset \prod_i \mathfrak{S}_{\mathbf{n}_i} = W_{\mathbf{n}}$. In terms of (4-3), the multiplication map $H^*_{T\times G_{\mathbf{d}}}(pt) \otimes H^*_{T\times G_{\mathbf{e}}}(pt) \to H^*_{T\times G_{\mathbf{n}}}(pt)$ now reads

$$(f * g)(\underline{z}_{[1,\mathbf{n}]}) = \sum_{\sigma \in Sh_{\mathbf{d},\mathbf{e}}} \sigma\left[K_{\mathbf{d},\mathbf{e}}(\underline{z}_{[1,\mathbf{n}]}) \cdot f(\underline{z}_{[1,\mathbf{d}]}) \cdot g(\underline{z}_{[\mathbf{d}+1,\mathbf{n}]})\right]$$

where $K_{\mathbf{d},\mathbf{e}}(\underline{z}_{[1,\mathbf{n}]}) = \prod_{s=0}^2 K^{(s)}_{\mathbf{d},\mathbf{e}}(\underline{z}_{[1,\mathbf{n}]})$ with

$$K^{(0)}_{\mathbf{d},\mathbf{e}}(\underline{z}_{[1,\mathbf{n}]}) = \prod_{i\in I} \prod_{\substack{1\leq l\leq \mathbf{d}_i \\ \mathbf{d}_i+1\leq k\leq \mathbf{n}_i}} (z_{i,l} - z_{i,k})^{-1},$$

$$K^{(1)}_{\mathbf{d},\mathbf{e}}(\underline{z}_{[1,\mathbf{n}]}) = \prod_{h\in\Omega} \prod_{\substack{1\leq l\leq \mathbf{d}_{h'} \\ \mathbf{d}_{h''}+1\leq k\leq \mathbf{n}_{h''}}} (z_{h',l} - z_{h'',k} - q_1) \prod_{\substack{1\leq l\leq \mathbf{d}_{h''} \\ \mathbf{d}_{h'}+1\leq k\leq \mathbf{n}_{h'}}} (z_{h'',l} - z_{h',k} - q_2)$$

[8] in other words $H_*^T(\Lambda_Q)$ and $H_*^T(T^*\mathfrak{M}_Q)$ are two different integral forms of the same $\mathbb{Q}(q_1, q_2)$-algebra; this explains the discrepancy between usual and nilpotent Kac polynomials.

and

$$K^{(2)}_{\mathbf{d},\mathbf{e}}(\underline{z}_{[1,\mathbf{n}]}) = \prod_{i\in I} \prod_{\substack{1\le l\le \mathbf{d}_i \\ \mathbf{d}_i+1\le k\le \mathbf{n}_i}} (z_{i,k} - z_{i,l} - q_1 - q_2).$$

As shown in Schiffmann and Vasserot [2017a], $H^T_*(T^*\mathfrak{M}_{Q,\mathbf{d}})$ and $H^T_*(\Lambda_{Q,\mathbf{d}})$ are torsion-free and of generic rank one as modules over $H^*_{T\times G_{\mathbf{d}}}(pt)$, hence the localization map i_* is injective and Theorem 4.4 yields a description of the cohomological Hall algebra $H^T_*(\Lambda_Q)$ as a subalgebra of the above shuffle algebra, generated by an explicit collection of polynomials. This allows one to identify the rational form of $H^T_*(\mathfrak{M}_Q)$ with the positive half of the Drinfeld Yangian $Y_h(\mathfrak{g}_Q)$ when Q is of finite type and with the positive half of the Yangian version of the elliptic Lie algebra $\mathfrak{g}_{Q_0}[s^{\pm 1}, t^{\pm 1}] \oplus K$ when Q is an affine Dynkin diagram. Here K is the full central extension of the double loop algebra $\mathfrak{g}_{Q_0}[s^{\pm 1}, t^{\pm 1}]$. Beyond these case, shuffle algebras tend to be rather difficult to study and the algebraic structure of $H^T_*(\mathfrak{M}_Q)$ (or $H^T_*(\Lambda_Q)$) is still mysterious (see, however Neguţ [2015, 2016] for some important applications to the geometry of instanton moduli spaces in the case of the Jordan or affine type A quivers).

There are analogous shuffle algebra models in the case of an arbitrary oriented Borel-Moore homology theory, but the torsion-freeness statement remains conjectural in general. In order to give the reader some idea of what these are, as well as for later use, let us describe the localized K-theoretic Hall algebra of the g-loop quiver S_g. In this case, there is a $g+1$-dimensional torus T_g acting by rescaling the arrows:

$$(\xi_1, \xi_2, \dots, \xi_g, p)\cdot(h_1, h_1^*, \dots, h_g, h_g^*) = (\xi_1 h_1, p\xi_1^{-1} h_1^*, \dots, \xi_g h_g, p\xi_g^{-1} h_g^*)$$

and we have an identification

$$Sh^{K^{T_g}}_{S_g} := \bigoplus_d K^{T_g\times G_d}(pt) \simeq \bigoplus_d \mathbb{Q}[\xi_1^{\pm 1}, \dots, \xi_g^{\pm 1}, p^{\pm 1}; z_1^{\pm 1}, \dots, z_d^{\pm 1}]^{\mathfrak{S}_d}.$$

The multiplication takes the form

$$(f * g)(\underline{z}_{[1,n]}) = \sum_{\sigma\in Sh_{d,e}} \sigma\Big[K_{d,e}(\underline{z}_{[1,n]})\cdot f(\underline{z}_{[1,d]})\cdot g(\underline{z}_{[d+1,n]})\Big] \tag{4-4}$$

where $K_{d,e}(\underline{z}_{[1,n]}) = \prod_{s=0}^2 K^{(s)}_{d,e}(\underline{z}_{[1,n]})$ with

$$K^{(0)}_{d,e}(\underline{z}_{[1,n]}) = \prod_{\substack{1\le l\le d \\ d+1\le k\le n}} (1 - z_l/z_k)^{-1}$$

$$K^{(1)}_{d,e}(\underline{z}_{[1,n]}) = \prod_{\substack{1\le l\le d \\ d+1\le k\le n}} \prod_{u=1}^{g} (1 - \xi_u^{-1} z_l/z_k)(1 - p^{-1}\xi_u z_l/z_k)$$

$$K^{(2)}_{d,e}(\underline{z}_{[1,n]}) = \prod_{\substack{1 \le l \le d \\ d+1 \le k \le n}} (1 - p^{-1} z_k / z_l)^{-1}.$$

4.4 PBW theorem. We finish this paragraph with the following important structural result due to Davison and Meinhardt:

Theorem 4.5 (Davison and Meinhardt [2016]). *There exists a graded algebra filtration $\mathbb{Q} = F_0 \subseteq F_1 \subseteq \cdots$ of $H_*(T^*\mathfrak{M}_Q)$ and an algebra isomorphism*

$$gr_{F_\bullet}(H_*(T^*\mathfrak{M}_Q)) \simeq Sym(\bar{\mathfrak{n}}_Q[u]) \tag{4-5}$$

where $\bar{\mathfrak{n}}_Q = \bigoplus_{\mathbf{d} \in \mathbb{N}^I} \bar{\mathfrak{n}}_{Q,\mathbf{d}}$ is a $\mathbb{N}^I \times \mathbb{N}$-graded vector space and $deg(u) = -2$. The same holds for $H_*(\Lambda_Q)$.

The filtration $F_\bullet$ is, essentially, the perverse filtration associated to the projection from the stack $T^*\mathfrak{M}_Q$ to its coarse moduli space. As a direct corollary, $F_1 \simeq \bar{\mathfrak{n}}_Q[u]$ is equipped with the structure of an $\mathbb{N}$-graded Lie algebra; it is easy to see that it is the polynomial current algebra of an $\mathbb{N}$-graded Lie algebra $\bar{\mathfrak{n}}_Q$ (coined the *BPS Lie algebra* in Davison and Meinhardt [ibid.]). Loosely speaking, Theorem 4.5 says that $H_*(T^*\mathfrak{M}_Q)$ is (a filtered deformation of) the enveloping algebra $U(\bar{\mathfrak{n}}_Q[u])$, i.e. some kind of Yangian of $\bar{\mathfrak{n}}_Q$. Comparing graded dimension and using Theorem 4.2 we deduce that for any $\mathbf{d} \in \mathbb{N}^I$

$$\dim_{\mathbb{Z}} \bar{\mathfrak{n}}_{Q,\mathbf{d}} = A_{Q,\mathbf{d}}(t). \tag{4-6}$$

Moreover, when Q has no edge loops, it can be shown that the degree zero Lie subalgebra $\bar{\mathfrak{n}}_Q[0]$ is isomorphic to the positive nilpotent subalgebra $\mathfrak{n}_Q$ of the Kac-Moody algebra $\mathfrak{g}_Q$. This implies at once both the positivity and the constant term conjectures for Kac polynomials, see Section 2.2 The same reasoning also yields a proof of the nilpotent versions of the Kac conjectures (for an arbitrary quiver) of Section 2.5, see Davison [2016].

At this point, the following conjecture appears inevitable:

Conjecture 4.6. *The graded Lie algebra $\bar{\mathfrak{n}}_Q$ is isomorphic to the positive subalgebra $\widetilde{\mathfrak{n}}_Q$ of $\widetilde{\mathfrak{g}}_Q$.*

Notice that the definition of $\widetilde{\mathfrak{g}}_Q$ involves the *usual* Hall algebra of Q (over all the finite fields $\mathbb{F}_q$), while that of $\bar{\mathfrak{n}}_Q$ involves the (two-dimensional) *cohomological* Hall algebra of Q and the complex geometry of $\mathrm{Rep}_{\mathbb{C}}\,\Pi_Q$. Conjecture 4.6 would follow from the fact that $\bar{\mathfrak{n}}_Q$ is the positive half of some graded Borcherds algebra $\bar{\mathfrak{g}}_Q$.

4.5 Action on Nakajima quiver varieties. One important feature of the cohomological Hall algebras $H^T_*(T^*\mathfrak{M}_Q)$, $H^T_*(\Lambda_Q)$ is that they act, via some natural correspondences, on the cohomology of Nakajima quiver varieties. Recall that the Nakajima quiver variety $\mathfrak{M}_Q(\mathbf{v},\mathbf{w})$ associated to a pair of dimension vectors $\mathbf{v},\mathbf{w}\in\mathbb{N}^I$ is a smooth quasi-projective symplectic variety, which comes with a proper morphism $\pi:\mathfrak{M}_Q(\mathbf{v},\mathbf{w})\to\mathfrak{M}_{Q,0}(\mathbf{v},\mathbf{w})$ to a (usually singular) affine variety. The morphism π is an example of a symplectic resolution of singularities, of which quiver varieties provide one of the main sources. The quiver variety also comes with a canonical (in general singular) lagrangian subvariety $\mathfrak{L}(\mathbf{v},\mathbf{w})$ which, when the quiver has no edge loops, is the central fiber of π. Examples include the Hilbert schemes of points on $\mathbb{A}^2$ or on Kleinian surfaces, the moduli spaces of instantons on theses same spaces, resolutions of Slodowy slices in nilpotent cones and many others (see e.g. Schiffmann [2008] for a survey of the theory of Nakajima quiver varieties). The following theorem was proved by Varagnolo [2000], based on earlier work by Nakajima in the context of equivariant K-theory (see Nakajima [1998, 2001]).

Theorem 4.7 (Varagnolo [2000]). *Let Q be without edge loops and $\mathbf{w}\in\mathbb{N}^I$. There is a geometric action of the Yangian $Y_h(\mathfrak{g}_Q)$ on $F_\mathbf{w}:=\bigoplus_\mathbf{v} H_*^{T\times G_\mathbf{w}}(\mathfrak{M}_Q(\mathbf{v},\mathbf{w}))$, preserving the subspace $L_\mathbf{w}:=\bigoplus_\mathbf{v} H_*^{T\times G_\mathbf{w}}(\mathfrak{L}_Q(\mathbf{v},\mathbf{w}))$. For Q of finite type, $L_\mathbf{w}$ is isomorphic to the universal standard $Y_h(\mathfrak{g}_Q)$-module of highest weight $\mathbf{w}$.*

In the above, the Yangian $Y_h(\mathfrak{g}_Q)$ is defined using Drinfeld's new realization, applied to an arbitrary Kac-Moody root system, see Varagnolo [ibid.]. Its precise algebraic structure is only known for Q of finite or affine type.

Theorem 4.8 (Schiffmann and Vasserot [2017a]). *For any Q and any $\mathbf{w}\in\mathbb{N}^I$ there is a geometric action of the cohomological Hall algebra $H_*^T(T^*\mathfrak{M}_Q)$ on $F_\mathbf{w}$. The diagonal action of $H_*^T(T^*\mathfrak{M}_Q)$ on $\prod_\mathbf{w} F_\mathbf{w}$ is faithful. The subalgebra $H_*^T(\Lambda_Q)$ preserves $L_\mathbf{w}$, which is a cyclic module.*

For quivers with no edge loops, we recover Varagnolo's Yangian action by Theorem 4.4; It follows that there exists a surjective map

$$Y_h^+(\mathfrak{g}_Q)\otimes\mathbb{Q}(q_1,q_2)\to H_*^T(T^*\mathfrak{M}_Q)\otimes\mathbb{Q}(q_1,q_2)$$

This map is an isomorphism for Q of finite or affine type. The action of $H_*^T(T^*\mathfrak{M}_Q)$ is constructed by means of general Hecke correspondences; in fact one can view $H_*^T(T^*\mathfrak{M}_Q)$ as the *largest* algebra acting on $F_\mathbf{w}$ via Hecke correspondences. Considering dual Hecke correspondences (or adjoint operators), one defines an opposite action of $H_*^T(T^*\mathfrak{M}_Q)$; it is natural to expect that these two actions extend to an action of some 'Drinfeld double' of $H_*^T(T^*\mathfrak{M}_Q)$, but this remains to be worked out. It is also natural to expect that $L_\mathbf{w}$ is a universal or standard module for $H_*^T(T^*\mathfrak{M}_Q)$, as is suggested by Hausel's formula for the Poincaré polynomial of $\mathfrak{M}_Q(\mathbf{v},\mathbf{w})$ or $\mathfrak{L}_Q(\mathbf{v},\mathbf{w})$, which involves the (full) Kac polynomials, see Hausel [2010] and Bozec, Schiffmann, and Vasserot [2017].

Theorem 4.8 has an obvious analog (with the same proof) for an arbitrary OBM theory (see Yang and Zhao [2014] for the construction of the action).

4.6 Relation to Maulik-Okounkov Yangians. We finish this section by very briefly mentioning yet another (conjectural) construction of the Lie algebra $\widetilde{\mathfrak{g}}_Q$, this time directly by means of the symplectic geometry of Nakajima quiver varieties. Using the theory of *stable enveloppes* for $\mathbb{C}^*$-actions on smooth symplectic varietes, Maulik and Okounkov constructed for any pair of dimension vectors $\mathbf{w}_1, \mathbf{w}_2$ a quantum R-matrix

$$R_{\mathbf{w},\mathbf{w}_2}(t) = 1 + \frac{\hbar}{v^{-1}}\mathbf{r}_{\mathbf{w}_1,\mathbf{w}_2} + O(v^{-2}) \in \mathrm{End}(F_{\mathbf{w}_1} \otimes F_{\mathbf{w}_2})[[v^{-1}]].$$

Applying the RTT formalism, we obtain a graded algebra $\mathbb{Y}_Q$ acting on all the spaces $F_{\mathbf{w}}$; similarly, from the classical R-matrices $\mathbf{r}_{\mathbf{w}_1,\mathbf{w}_2}$ we obtain a $\mathbb{Z}$-graded Lie algebra $\mathbf{g}_Q$. Moreover, $\mathbb{Y}_Q$ is (up to some central elements) a filtered deformation of $U(\mathbf{g}_Q[u])$ (hence the name *Maulik-Okounkov Yangian* see Maulik and Okounkov [2012]). It can be shown that $\mathbf{g}_Q$ is a graded Borcherds algebra. The following conjecture was voiced by Okounkov [n.d.]:

Conjecture 4.9 (Okounkov)**.** *For any Q and any $\mathbf{d} \in \mathbb{N}^I$ we have* $\dim_{\mathbb{Z}}\ \mathbf{g}_Q[\mathbf{d}] = A_{Q,\mathbf{d}}(t)$.

The conjecture is known when Q is of finite type. Comparing with (3-4) and (4-6) leads to the following

Conjecture 4.10. *For any Q we have* $\mathbf{g}_Q \simeq \bar{\mathfrak{g}}_Q \simeq \widetilde{\mathfrak{g}}_Q$.

As a first step towards the above conjecture, we have

Theorem 4.11 (Schiffmann and Vasserot [2017b])**.** *For any Q there is a canonical embedding $H^T_*(\Lambda_Q) \to \mathbb{Y}^+_Q$, compatible with the respective actions on $\prod_{\mathbf{w}} F_{\mathbf{w}}$ of $H^T_*(\Lambda_Q)$ and $\mathbb{Y}_Q$.*

All together, we see that there are conjecturally (at least) *three* different incarnations of the *same* graded Borcherds Lie algebra: as a 'generic form' of the full Hall algebra of the category of representations of Q over finite fields, as a cohomological Hall algebra of the complex (singular) stack $T^*\mathfrak{M}_Q$, or as an algebra acting on the cohomology of Nakajima quiver varieties via R-matrices constructed by means of symplectic geometry. The precise structure of this graded Borcherds algebra, in particular its Cartan matrix (determined by the polynomials $C^{\mathrm{abs}}_{Q,\mathbf{d}}(t)$ counting the dimensions of the spaces of 'absolutely' cuspidal functions for Q) remains however very mysterious. From the symplectic geometry perspective, one might expect the polynomials $C^{\mathrm{abs}}_{Q}(t)$ to be related to some motive inside the Nakajima quiver variety, but there is no conjectural construction of such a motive that we know of.

In the remainder of this paper, we shift gears and consider categories of coherent sheaves on smooth projective curves instead of representations of quivers; motivated by the analogy with quivers, we will describe the spherical Hall algebra, full Hall algebra and $2d$-cohomological Hall algebras (!) of a curve, as well as a geometric interpretation of Kac polynomials in terms of Higgs bundles. We finish with some speculation about the Lie theoretic structures that we believe are lurking in the background.

5 Hall algebras of curves and shuffle algebras

5.1 Notations. We fix an integer $g \geq 0$ and a smooth, geometrically connected, projective curve X of genus g over a field $\mathbf{k}$. We denote by $\mathrm{Coh}(X)$ the category of coherent sheaves on X and by $\mathfrak{M}_{X,r,d}$ the stack of coherent sheaves of rank r and degree d. It is a smooth stack, locally of finite type. The Euler form is given by the Riemann-Roch formula

$$\langle \mathcal{F}, \mathcal{G} \rangle = (1-g)rk(\mathcal{F})rk(\mathcal{G}) + (rk(\mathcal{F})deg(\mathcal{G}) - rk(\mathcal{G})deg(\mathcal{F})).$$

Set $\mathfrak{M}_X = \bigsqcup_{r,d} \mathfrak{M}_{X,r,d}$.

5.2 Ringel–Hall algebra of a curve. Let us now assume that $\mathbf{k}$ is a finite field, and define

$$\mathbf{H}_X = \bigoplus_{r,d} \mathbf{H}_X[r,d], \qquad \mathbf{H}_X[r,d] := \mathrm{Fun}(\mathfrak{M}_{X,r,d}(\mathbf{k}), \mathbb{C}).$$

In an entirely similar fashion to (2-2), there is a convolution diagram

$$\mathfrak{M}_X \times \mathfrak{M}_X \xleftarrow{\;q\;} \widetilde{\mathfrak{M}}_X \xrightarrow{\;p\;} \mathfrak{M}_X \tag{5-1}$$

where $\widetilde{\mathfrak{M}}_X$ is the stack parametrizing short exact sequences

$$0 \longrightarrow \mathcal{F} \longrightarrow \mathcal{H} \longrightarrow \mathcal{G} \longrightarrow 0$$

in $\mathrm{Coh}(X)$; The map p is proper, while the map q is again a stack vector bundle, whose restriction to $\mathfrak{M}_{X,r,d} \times \mathfrak{M}_{X,r',d'}$ is of rank $-\langle (r',d'),(r,d)\rangle$. Setting $v = (\#\mathbf{k})^{\frac{1}{2}}$, $\mathbf{K} = \mathbb{C}[k_{(0,1)}^{\pm 1}, k_{(1,0)}^{\pm 1}]$ we equip, using (2-3) and (2-4), $\widetilde{\mathbf{H}}_X := \mathbf{H}_X \otimes \mathbf{K}$ with the structure of a $\mathbb{Z}^2$-graded bialgebra called the *Hall algebra* of X. It carries a nondegenerate Hopf pairing by (2-5). Finally, $\mathbf{H}_X \simeq \mathbf{H}_X^{\mathrm{bun}} \ltimes \mathbf{H}_X^0$ where $\mathbf{H}_X^{\mathrm{bun}}, \mathbf{H}_X^0$ are the Hall subalgebras of vector bundles, resp. torsion sheaves. The Hall algebra of a curve was first considered by M. M. Kapranov [1997], who established the (direct) dictionary between $\mathbf{H}_X^{\mathrm{bun}}$ and the space of automorphic forms[9] for the groups $GL(n, \mathbb{A}_X)$, together with the operations of parabolic induction (Eisenstein series) and restriction (constant term map), and who observed a striking analogy between $\widetilde{\mathbf{H}}_X$ and quantum loop groups.

Let $\mathbf{H}_X^{\mathrm{sph}}$ (resp. $\mathbf{H}_X^{\mathrm{sph,bun}}$) be the subalgebra of $\mathbf{H}_X$ generated by the constant functions on $\mathfrak{M}_{X,r,d}$ for $r = 0, 1$ (resp. $r = 1$) and $d \in \mathbb{Z}$. From the point of view of automorphic forms, $\mathbf{H}_X^{\mathrm{sph,bun}}$ is the space spanned by all components of Eisenstein series induced from trivial (i.e constant) automorphic forms for the torus (together with a suitable space of Hecke operators in the case of $\mathbf{H}_X^{\mathrm{sph}}$). One can show that $\widetilde{\mathbf{H}}_X^{\mathrm{sph}} := \mathbf{H}_X^{\mathrm{sph}} \otimes \mathbf{K}$ is a self-dual sub-Hopf algebra of $\widetilde{\mathbf{H}}_X$ which contains the characteristic functions of any Harder-Narasimhan strata (see Schiffmann [2011]). Moreover, contrary to $\widetilde{\mathbf{H}}_X$ which depends strongly on the fine arithmetic structure of X, $\widetilde{\mathbf{H}}_X^{\mathrm{sph}}$ only depends on the Weil numbers of X and admits an R_g*-rational form*, i.e. there exists a torsion-free R_g-Hopf algebra $\widetilde{\mathbf{H}}_{\Sigma_g}^{\mathrm{sph}}$ such that for any smooth projective curve X defined over a finite field, $\widetilde{\mathbf{H}}_{\Sigma_g}^{\mathrm{sph}} \otimes_{R_g} \mathbb{C}_X \simeq \widetilde{\mathbf{H}}_X^{\mathrm{sph}}$, where $\mathbb{C}_X$ is the R_g-module corresponding to the evaluation morphism $R_g \to \overline{\mathbb{Q}_l} \simeq \mathbb{C}$, $f \mapsto f(Fr_x)$. We view $\widetilde{\mathbf{H}}_{\Sigma_g}^{\mathrm{sph}}$ as some kind of (half) quantum group which depends on $\dim T_g = g + 1$ quantum parameters, associated to curves of genus g. The full quantum group is, as before, obtained by the Drinfeld double procedure. Hall algebras in genus 0 and 1 are already very interesting:

Theorem 5.1 (M. M. Kapranov [1997],Baumann and Kassel [2001])**.** *The Drinfeld double* $\mathbf{D}\widetilde{\mathbf{H}}_{\Sigma_0}^{sph}$ *is isomorphic to the quantum affine algebra* $U_v(\widehat{\mathfrak{sl}_2})$.

Theorem 5.2 (Burban and Schiffmann [2012],Schiffmann and Vasserot [2011])**.** *The Drinfeld double* $\mathbf{D}\widetilde{\mathbf{H}}_{\Sigma_1}^{sph}$ *is isomorphic to the spherical double affine Hecke algebra* $\mathbf{S}\ddot{\mathbf{H}}_{q,t}(GL_\infty)$ *of type* $GL(\infty)$.

In the above two cases, the structure of the Hall algebra $\widetilde{\mathbf{H}}_{\Sigma_g}^{\mathrm{sph}}$ is rather well understood: it has a PBW-type basis as well as a canonical basis constructed from simple perverse sheaves (Eisenstein sheaves) on the stacks $\mathfrak{M}_X$. The spherical Hall algebra $\widetilde{\mathbf{H}}_{\Sigma_1}^{\mathrm{sph}}$ –also called the *elliptic Hall algebra*– has found a surprising number of applications in representation theory of Cherednik algebras, low-dimensional topology and knot theory (e.g. Morton and Samuelson [2017] , Gorsky and Neguţ [2015]), algebraic geometry and mathematical physics of the instanton spaces on $\mathbb{A}^2$ (e.g. Schiffmann and Vasserot [2013b], Schiffmann and Vasserot [2013a], and Neguţ [2016]), combinatorics of Macdonald polynomials (e.g. Bergeron, Garsia, Leven, and Xin [2016] and Francesco and Kedem [2017]),

[9] one nice feature of the Hall algebra is that it incorporates the algebra of Hecke operators, as the sub Hopf algebra of torsion sheaves $\mathbf{H}_X^0$

categorification (e.g. Cautis, Lauda, Licata, Samuelson, and Sussan [2016]), etc. The elliptic Hall algebra (or close variants thereof) has independently appeared in the work of Miki [2007] and Ding and Iohara [1997] and Feigin and his collaborators (see e.g., B. Feigin, E. Feigin, Jimbo, Miwa, and Mukhin [2011]).

5.3 Shuffle algebra presentation. Although the structure of $\widetilde{\mathbf{H}}^{\mathrm{sph}}_{\Sigma_g}$ for $g > 1$ is much less well understood, we always have a purely algebraic model of $\widetilde{\mathbf{H}}^{\mathrm{sph}}_{\Sigma_g}$ at our disposal, once again[10] in the guise of a shuffle algebra. More precisely, let

$$\zeta_{\Sigma_g}(z) := \frac{\prod_{i=1}^{g}(1-\eta_i z)}{(1-z)(1-qz)} \in R_g(z)$$

be the 'generic' zeta function of a curve of genus g and put $\zeta'_{\Sigma_g}(z) = (1-qz)(1-qz^{-1})\zeta_{\Sigma_g}(z)$. Consider the shuffle algebra

$$Sh_{\Sigma_g} := \bigoplus_d R_g[z_1^{\pm 1}, \ldots, z_d^{\pm 1}]^{\mathfrak{S}_d}$$

with multiplication

$$(f * g)(\underline{z}_{[1,n]}) = \sum_{\sigma \in Sh_{d,e}} \sigma\left[K_{d,e}(\underline{z}_{[1,n]}) f(\underline{z}_{[1,d]}) g(\underline{z}_{[d+1,n]})\right] \tag{5-2}$$

where $K_{d,e}(\underline{z}_{[1,n]}) = \prod_{1 \leq l \leq d < k \leq n} \zeta'_{\Sigma_g}(z_l/z_k)$, for any d, e and $n = d+e$.

Theorem 5.3 (Schiffmann and Vasserot [2012]). *The assignment* $1_{Pic^d} \mapsto z^d \in Sh_{\Sigma_g}[1]$ *extends to an* R_g*-algebra embedding* $\Psi : \mathbf{H}^{sph,bun}_{\Sigma_g} \to Sh_{\Sigma_g}$.

The map Ψ is essentially the iterated coproduct $\Delta^{(r)} : \mathbf{H}^{\mathrm{sph}}_{\Sigma_g}[r] \to (\mathbf{H}^{\mathrm{sph}}_{\Sigma_g}[1])^{\otimes r}$. In the language of automorphic forms, Theorem 5.3 amounts to the Langlands formula for the constant term of Eisenstein series, or equivalently to the Gindikin-Karpelevich formula. One should not be deceived by the apparent simplicity of the shuffle algebra description for $\mathbf{H}^{\mathrm{sph,bun}}_X$. In particular, $\mathbf{H}^{\mathrm{sph,bun}}_{\Sigma_g}$ is *not* free over R_g and the relations satisfied by the generators 1_{Pic^d} of $\mathbf{H}^{\mathrm{sph,bun}}_X$ *do* depend on the arithmetic of the Weil numbers $\sigma_1, \ldots, \sigma_{2g}$ of X (more precisely, on the $\mathbb{Z}$-linear dependence relation between $log(\sigma_1), \ldots log(\sigma_{2g})$, the so-called *wheel relations*).

[10] There is nothing surprising in the ubiquity of shuffle algebras: any finitely generated, $\mathbb{N}$-graded self-dual Hopf algebra has a realization as a shuffle algebra

5.4 Geometric Langlands isomorphism. Comparing (4-4) and (5-2) we immediately see that, up to the identification $p = q^{-1}, \xi_i = \eta_i^{-1}$ we have $Sh_{\Sigma_g} \simeq Sh_{S_g}^{K^{T_g}}$. This implies that there is an isomorphism

$$\Phi : \mathbf{H}_{\Sigma_g}^{\mathrm{sph,bun}} \xrightarrow{\sim} \mathbf{K}_{S_g}^{\mathrm{sph},T_g} \tag{5-3}$$

where $\mathbf{K}_{S_g}^{\mathrm{sph},T_g}$ is the subalgebra of $K^{T_g}(T^*\mathfrak{M}_{S_g})$ generated by its rank one component $K^{T_g}(T^*\mathfrak{M}_{S_g,1})$. The isomorphism (5-3) between the (spherical) Hall algebra of X and the (spherical) K-theoretical Hall algebra of the S_g quiver should be viewed as an incarnation, at the level of Grothendieck groups, of a (linearized) form of geometric Langlands correspondence. Indeed, the Langlands philosophy predicts an equivalence

$$Coh(LocSys_r(X_{\mathbb{C}})) \simeq D\text{–}mod(Bun_r(X_{\mathbb{C}}))$$

between suitable (infinity) categories of coherent sheaves on the moduli stack of GL_r-local systems on $X_{\mathbb{C}}$ and D-modules on the stack of GL_r-bundles on $X_{\mathbb{C}}$. On one hand, the formal neighborhood in $LocSys_r(X_{\mathbb{C}})$ of the trivial local system may be linearized as the formal neighborhood of 0 in $\{(x_1,\dots,x_g,y_1,\dots,y_g) \in \mathfrak{gl}_r(\mathbb{C})^{2g} \mid \sum_i [x_i,y_i] = 0\}/GL_r(\mathbb{C}) \simeq T^*\mathfrak{M}_{S_g,r}$; on the other hand, $\mathbf{H}_{\Sigma_g}^{sph,bun}$ may be lifted to a category of holonomic D-modules (perverse sheaves) on $Bun_r(X_{\mathbb{C}})$. We refer to Schiffmann and Vasserot [2012] for a more detailed discussion.

5.5 Variations. There are several interesting variants of the above constructions and results: one can consider categories of D-parabolic coherent sheaves (see Schiffmann [2004] and Lin [2014]); this yields, for instance, some quantum affine or toroidal algebras. One can also consider arithmetic analogs the Hall algebra, replacing the abelian category of coherent sheaves on a curve X over a finite field by coherent sheaves or vector bundles (in the sense of Arakelov geometry) over $\overline{Spec(\mathcal{O}_K)}$, where K is a number field. The case of $K = \mathbb{Q}$ is discussed in M. Kapranov, Schiffmann, and Vasserot [2014], where the spherical Hall algebra is described as some analytic shuffle algebra with kernel given by the Riemann zeta function $\zeta(z)$.

6 Counting Cuspidals and Cohomological Hall algebras of curves

6.1 Counting cuspidals and the full Hall algebra of a curve. In Sections 5.2 and 5.3 we considered the spherical Hall algebra $\mathbf{H}_X^{\mathrm{sph}}$ for which we provided a shuffle algebra description involving the zeta function $\zeta_X(z)$. What about the whole Hall algebra $\mathbf{H}_X$? As shown in M. Kapranov, Schiffmann, and Vasserot [2017], $\mathbf{H}_X^{bun}$ admits a shuffle description as well, but it is much less explicit than for $\mathbf{H}_X^{\mathrm{sph}}$. Recall that an element $f \in \mathbf{H}_X^{\mathrm{bun}}[r,d]$ is *cuspidal* if it is quasi-primitive, i.e. if

$$\Delta(f) \in f \otimes 1 + k_{r,d} \otimes f + \mathbf{H}_X^0 \otimes \widetilde{\mathbf{H}}_X[r].$$

The algebra $\mathbf{H}_X^{\mathrm{bun}}$ is generated by the spaces of cuspidal elements

$$\mathbf{H}_X^{\mathrm{cusp}} = \bigoplus_{r,d} \mathbf{H}_X^{\mathrm{cusp}}[r,d]$$

and $\dim \mathbf{H}_X^{\mathrm{cusp}}[r,d] < \infty$ for all r,d. The function field Langlands program (Lafforgue [2002]) sets up a correspondence $\chi \mapsto f_\chi$ between characters $\chi : \mathbf{H}_X^0 \to \mathbb{C}$ associated to rank r irreducible local systems on X and cuspidal Hecke eigenfunctions $f_\chi = \sum_d f_{\chi,d} \in \prod_d \mathbf{H}_X^{\mathrm{cusp}}[r,d]$. The shuffle algebra description of $\mathbf{H}_X$ is as follows: we have a family of variables $z_{\chi,i}$, $i \in \mathbb{N}$ for *each* cuspidal Hecke eigenform f_χ (up to $\mathbb{G}_m$-twist) and the shuffle kernels involve the Rankin-Selberg L-functions $L(\chi,\chi',z)$ of pairs of characters χ,χ' in place of $\zeta_X(z)$ (see Fratila [2013] for a full treatment when $g=1$). In principle, one could try, using the PBW theorem and arguing as in the case of quivers (see Section 3.2), to deduce from the above shuffle description of $\mathbf{H}_X$ an expression for the dimensions of the spaces of cuspidal functions $\mathbf{H}_X^{\mathrm{cusp}}[r,d]$ (or better, absolutely cuspidal functions $\mathbf{H}_X^{\mathrm{abs.\ cusp}}[r,d]$) in terms of the Kac polynomials $A_{g,r}$. Very recently H. Yu managed –by other means[11]– to compute the dimension of $\mathbf{H}_X^{\mathrm{abs.\ cusp}}[r,d]$ directly:

Theorem 6.1 (Yu [n.d.]). *For any g,r there exists a (unique) polynomial $C_{g,r}^{abs} \in R_g$ such that for any smooth projective curve X of genus g defined over a finite field, $\dim \mathbf{H}_X^{abs.\ cusp}[r,d] = C_{g,r}^{abs}(Fr_X)$.*

This generalizes a famous result of Drinfeld (for $r=2$, Drinfel'd [1981]) and proves a conjecture of Deligne [2015] and Kontsevich [2009]. Interestingly, the polynomial $C_{g,r}^{\mathrm{abs}}$ is explicit: Yu expresses it in terms of the numbers of rational points of the moduli spaces of stable Higgs bundles over finite fields and hence, by Theorem 7.1 below, in terms of the Kac polynomials $A_{g,r}$! This strengthens our belief that the structure of $\mathbf{H}_X$ as an associative algebra is nice enough that it should have a character formula similar to that of Borcherds algebras. For instance, we have

$$A_{g,1}(F) = C_{g,1}^{\mathrm{abs}}(F), \qquad A_{g,2}(F) = C_{g,2}^{\mathrm{abs}}(F) + (g-1)C_{g,1}^{\mathrm{abs}}(F)^2 + C_{g,1}^{\mathrm{abs}}(F)$$

$$\begin{aligned} A_{g,3}(F) = & C_{g,3}^{\mathrm{abs}}(F) + (g-1)C_{g,1}^{\mathrm{abs}}(F)\left\{4C_{g,2}^{\mathrm{abs}}(F) + C_{g,1}^{\mathrm{abs}}(F^2) + 2(g-1)C_{g,1}^{\mathrm{abs}}(F)^2\right\} \\ & + 4(g-1)C_{g,1}^{\mathrm{abs}}(F)^2 + C_{g,1}^{\mathrm{abs}}(F) \end{aligned}$$

where $F = Fr_X$.

[11] namely, using the Arthur-Selberg trace formula

Conjecture 6.2. *For any g, r there exists a (non-virtual) $GSp(2g, \overline{\mathbb{Q}_l})$-representation $\mathbb{C}^{abs}_{g,r}$ such that $C^{abs}_{g,r} = \tau(ch(\mathbb{C}^{abs}_{g,r}))$.*

6.2 Cohomological Hall algebra of Higgs sheaves. We take $\mathbf{k} = \mathbb{C}$ here. The (underived) cotangent stack $T^*\mathfrak{M}_X = \bigsqcup_{r,d} T^*\mathfrak{M}_{X,r,d}$ is identified with the stack of Higgs sheaves $\mathcal{H}iggs_X = \bigsqcup_{r,d} \mathcal{H}iggs_{X,r,d}$, which parametrizes pairs $(\mathcal{F}, \theta)$ with $\mathcal{F} \in \mathrm{Coh}(X)$ and $\theta \in Hom_{\mathcal{O}_X}(\mathcal{F}, \mathcal{F} \otimes \Omega_X)$. The *global nilpotent cone* $\Lambda_X = \bigsqcup_{r,d} \Lambda_{X,r,d}$ is the closed Lagrangian substack whose objects are Higgs sheaves $(\mathcal{F}, \theta)$ for which θ is nilpotent. Both stacks are singular, locally of finite type and have infinitely many irreducible components. The stack $\Lambda_{X,r,d}$ is slightly better behaved since all irreducible components are of dimension $(g-1)r^2$. We refer to Bozec [2017] for an explicit description of these irreducible components. The torus $T = \mathbb{G}_m$ acts on $\mathcal{H}iggs_X$ and Λ_X by $t \cdot (\mathcal{F}, \theta) = (\mathcal{F}, t\theta)$.

Theorem 6.3 (Sala and Schiffmann [2018] and Minets [2018])**.** *The Borel-Moore homology spaces $H^T_*(\mathcal{H}iggs_X)$ and $H^T_*(\Lambda_X)$ carry natural $\mathbb{Z} \times \mathbb{Z}^2$-graded associative algebra structures. Moreover, the direct image morphism*

$$i_* : H^T_*(\Lambda_X) \to H^T_*(\mathcal{H}iggs_X)$$

is an algebra homomorphism.

The definition of the algebra structure roughly follows the same strategy as for quivers: working in local charts, we use the construction of $\mathcal{H}iggs_X$ as a symplectic reduction to embed everything into some smooth moduli stacks. As before, the morphism i_* becomes invertible after localizing with respect to $H^*_T(pt)$. There is an embedding of $\mathfrak{M}_X$ in Λ_X as the zero section of the projection $p : \mathcal{H}iggs_X \to \mathfrak{M}_X$; its image is an irreducible component of Λ_X.

Theorem 6.4 (Sala and Schiffmann [2018])**.** *The algebra $H^T_*(\Lambda_X)$ is generated by the collection of subspaces $H^T_*(\mathfrak{M}_{X,r,d})$ for $(r, d) \in \mathbb{Z}^2$.*

What about a shuffle algebra description of $H^T_*(\Lambda_X)$? The cohomology ring $\mathbb{H}_{r,d} = H^*(\mathfrak{M}_{X,r,d})$ acts on $H^T_*(\Lambda_{X,r,d})$ by $c \cdot h = p^*(c) \cap h$. By Heinloth's extension of the Atiyah-Bott theorem Heinloth [2012], $\mathbb{H}_{0,d} = S^d(H^*(X)[z])$ for $d \geq 1$ while for $r \geq 1, d \in \mathbb{Z}$, $\mathbb{H}_{r,d} = \mathbb{Q}[c_{i,\gamma}(\mathcal{E}_{r,d}) \mid i \geq 1, \gamma]$ is a polynomial algebra in the Künneth components $c_{i,\gamma}(\mathcal{E}_{r,d})$ of the Chern classes of the tautological sheaf $\mathcal{E}_{r,d}$ over $\mathfrak{M}_{X,r,d} \times X$; here γ runs over a basis $\{1, a_1, b_1, \ldots, a_g, b_g, \varpi\}$ of $H^*(X)$.

Theorem 6.5 (Sala and Schiffmann [2018] and Minets [2018])**.** *The $\mathbb{H}_{r,d}$-module $H^T_*(\Lambda_{X,r,d})$ is torsion-free and of generic rank one.*

This opens up the possibility to construct a (shuffle) algebra structure on $\bigoplus_{r,d} \mathbb{H}_{r,d}$, but this has so far only be achieved for the subalgebra of torsion sheaves, see Minets [2018]. Very similar shuffle algebras occur, not surprisingly, as operators on the cohomology of moduli spaces of semistable sheaves on smooth surfaces, see Neguţ [2017].

7 Kac polynomials and Poincaré polynomials of moduli of stable Higgs bundles

7.1 Moduli spaces of stable Higgs bundles. Recall that a Higgs sheaf (V, θ) on X is *semistable* if for any subsheaf $W \subset V$ such that $\theta(W) \subseteq W \otimes \Omega_X$ we have $\mu(W) \leq \mu(V)$, where $\mu(\mathcal{F}) = deg(\mathcal{F})/rk(\mathcal{F})$ is the usual slope function. Replacing $\leq$ by $<$ we obtain the definition of a *stable* Higgs sheaf. The open substack $\mathcal{H}iggs^{st}_{X,r,d} \subset \mathcal{H}iggs_{X,r,d}$ of stable sheaves is a $\mathbb{G}_m$-gerbe over a smooth quasi-projective symplectic variety $Higgs^{st}_{r,d}$. On the contrary, the open substack $\mathcal{H}iggs^{ss}_{r,d} \subset \mathcal{H}iggs_{X,r,d}$ of semistable sheaves is singular as soon as $gcd(r,d) > 1$; when $gcd(r,d) = 1$, $\mathcal{H}iggs^{ss}_{X,r,d} = \mathcal{H}iggs^{st}_{X,r,d}$. The variety $Higgs^{st}_{X,r,d}$ has played a fundamental role in algebraic geometry, in the theory of integrable systems, in the geometric Langlands program, in the theory of automorphic forms, and is still the focus of intensive research; we refer to e.g. Hausel [2013] for a survey of the many important conjectures in the subject.

7.2 Poincaré polynomials and Kac polynomials. Assume that $\mathbf{k} = \mathbb{F}_q$. The following result provides a very vivid geometric interpretation of the Kac polynomial $A_{g,r}$.

Theorem 7.1 (Schiffmann [2016] and Mozgovoy and Schiffmann [2017]). *For any r, d with $gcd(r,d) = 1$ we have*

$$\#Higgs^{st}_{X,r,d}(\mathbb{F}_q) = q^{(g-1)r^2+1} A_{g,r}(Fr_X).$$

Together with some purity argument and Mellit's simplification of the explicit formula for $A_{g,r}$, this proves the conjecture of Hausel and Rodriguez-Villegas [2008] for the Poincaré polynomial of $Higgs^{st}_{X,r,d}$ (for $\mathbf{k} = \mathbb{C}$ or $\mathbb{F}_q$). There are two proofs of Theorem 7.1: one is based on a deformation argument to relate directly the point count of $Higgs_{X,r,d}(\mathbb{F}_q)$ to the count of indecomposable sheaves; the other uses Hall-theoretic techniques (for the category of positive Higgs sheaves). Both have been generalized or put in a broader context, see Dobrovolska, Ginzburg, and Travkin [2016] and Fedorov, A. Soibelman, and Y. Soibelman [2017].

Concerning the stable global nilpotent cone $\Lambda^{st}_{X,r,d} = \Lambda_{X,r,d} \cap Higgs^{st}_{X,r,d}$ we have the following interpretation of the constant term of the Kac polynomial:

Theorem 7.2 (Schiffmann [2016]). *For any* (r,d) *with* $gcd(r,d)=1$ *we have*

$$\#Irr(\Lambda^{st}_{X,r,d}) = A_{g,r,d}(0)$$

Recall that by Proposition 1.7 we have $A_{g,r}(0) = A_{S_g,r}(1)$; this suggests the existence of some natural partition of $Irr(\Lambda^{st}_{X,r,d})$, but the geometric meaning of such a partition is unclear to us.

7.3 Donaldson-Thomas invariants and Kac polynomials. What about *non* coprime (r,d) ? In this case, it is still possible to perform the (orbifold) point count of the stack $\mathcal{H}iggs_{X,r,d}(\mathbf{k})$ when $\mathbf{k}=\mathbb{F}_q$; this point count is best expressed in terms of the Donaldson-Thomas invariants $\Omega_{X,r,d}$ which are defined by the following generating series:

$$\forall\ \nu\in\mathbb{Q},\qquad \sum_{\frac{d}{r}=\nu}\frac{\Omega_{X,r,d}}{q-1}w^r z^d := \mathrm{Log}\left(\sum_{\frac{d}{r}=\nu} q^{(1-g)r^2}\#(\mathcal{H}iggs^{ss}_{X,r,d}(\mathbb{F}_q))w^r z^d\right).$$

Theorem 7.3 (Mozgovoy and Schiffmann [2017]). *For any* r,d *we have* $\Omega_{X,r,d} = qA_{g,r}(Fr_X)$.

If $gcd(r,d)=1$ then

$$\Omega_{X,r,d} = (q-1)q^{(1-g)r^2}\#(\mathcal{H}iggs^{st}_{X,r,d}(\mathbb{F}_q)) = q^{(1-g)r^2}\#Higgs^{st}_{X,r,d}(\mathbb{F}_q)$$

so that we recover Theorem 7.1.

8 Delirium Tremens: a hierarchy of Lie algebras

We conclude this survey with some wild speculations concerning potential Lie algebras associated to curves, rather than to quivers

8.1 Lie algebras from curves ? Following the analogy with quivers, it is natural to expect the existence of a family of $\mathbb{Z}^2$-graded complex Lie algebras $\mathfrak{g}_g = \bigoplus_{r,d}\mathfrak{g}_g[r,d]$ such that $\mathbf{H}^{\mathrm{sph}}_g$ is a $(g+1)$-quantum parameter deformation of $U^+(\mathfrak{g}_g)$, and

$$\dim\ \mathfrak{g}_g[r,d] = A_{g,r}(0,\ldots,0)$$

for any r,d. This Lie algebra $\mathfrak{g}_g$ would be a curve analog of the Kac-Moody algebra $\mathfrak{g}_Q$ associated to a quiver Q (or its variant $\mathfrak{g}^B_Q$ if Q has edge loops). What about the the analog of the graded Borcherds algebra $\widetilde{\mathfrak{g}}_Q$? Because the grading in the context of curves is by the character ring of $GSp(2g,\overline{\mathbb{Q}_l})$ rather than by $\mathbb{Z}$, it seems natural to expect the existence of a Lie algebra $\widetilde{\mathfrak{g}}_g$ *in the tensor category of finite-dimensional* $GSp(2g, \overline{\mathbb{Q}_l})$*-modules*, with $\mathfrak{g}_g$ being identified with the sub-Lie algebra corresponding to the tensor subcategory of trivial representations (of arbitrary rank). Moreover, we should have

$$\widetilde{\mathfrak{g}}_g[r,d] = \mathbb{A}_{g,r} \in GSp(2g, \overline{\mathbb{Q}_l}) - \text{mod},$$

where $\mathbb{A}_{g,r}$ is as in Conjecture 1.6, and the cuspidal (or 'simple root vectors') of $\widetilde{\mathfrak{g}}_g[r,d]$ should form a subrepresentation isomorphic to $\mathbb{C}^{\text{abs}}_{g,r}$, where $\mathbb{C}^{\text{abs}}_{g,r}$ is as in Conjecture 6.2.

Although we do not have any clue at the moment as to what $\widetilde{\mathfrak{g}}_g$ could be, the Langlands isomorphism (5-3) provides us with a very good guess concerning $\mathfrak{g}_g$. Namely, it is expected that the (spherical) K-theoretical Hall algebra $K^T(T^*\mathfrak{M}_Q)$– by analogy with the (spherical) cohomological Hall algebra $H^T_*(T^*\mathfrak{M}_Q)$– is a deformation of $U^+(\widetilde{\mathfrak{g}}_Q[u^{\pm 1}])$. This strongly suggests that, at least as a vector space, $\mathfrak{g}_g \simeq \widetilde{\mathfrak{g}}_{S_g}[u^{\pm 1}]$. Note that the $\mathbb{N}$-grading of $\widetilde{\mathfrak{g}}_{S_g}$ gets lost in the process since there is no obvious grading in the K-theoretical Hall algebra. Taking graded dimensions, we obtain the equality $A_{S_g,r}(1) = A_{g,r}(0,\ldots,0)$ of Proposition 1.7. Of course, this is not a proof but rather a conceptual explanation of this equality. We summarize this in the chain of inclusions of Lie algebras

$$\mathfrak{g}_{S_g} \subseteq \tilde{\mathfrak{g}}_{S_g} \subset \tilde{\mathfrak{g}}_{S_g}[t,t^{-1}] \simeq \mathfrak{g}_g \subset \tilde{\mathfrak{g}}_g \subset \tilde{\mathfrak{g}}_g[t,t^{-1}].$$

Examples. i) Suppose $g = 0$. Then $\widetilde{\mathfrak{g}}_{S_0} = \mathfrak{g}_{S_0} = \mathfrak{sl}_2$ and $\mathfrak{g}_0 \simeq \widehat{\mathfrak{sl}_2}$, while it is natural to expect that $\widetilde{\mathfrak{g}}_0 = (\mathfrak{sl}_2 \oplus K)[u^{\pm 1}]$, where K is a one dimensional central extension, placed in degree one. Note that we have $A_{0,0} = q+1$, $A_{0,1} = 1$ and $A_{0,r} = 0$ for $r > 1$.
ii) Suppose $g = 1$. Then $\widetilde{\mathfrak{g}}_{S_1} \simeq \mathfrak{g}_{S_1} = \overline{\mathbb{Q}_l}[s^{\pm 1}]$ is the Heisenberg algebra, and $\mathfrak{g}_1 = \overline{\mathbb{Q}_l}[s^{\pm 1}, t^{\pm 1}] \oplus K_1 \oplus K_2$, where $K_1 \oplus K_2$ is a two-dimensional central extension. The Lie algebra structure is not the obvious one however, but rather a central extension of the Lie bracket $[s^r t^d, s^n t^m] = (rm - dn)s^{r+n}t^{d+m}$ (see Schiffmann and Vasserot [2013a], App. F. for the case of the Yangian).

8.2 Summary of Hall algebras, their corresponding Lie algebras and Kac polynomials. We conclude this survey with the following table, containing our heuristics.

Type of Hall Algebras	Quivers	Curves
K-thr. Hall algebra $K^{sph,T}(T^*\mathfrak{M})$	$\widetilde{\mathfrak{g}}_Q[u^{\pm 1}]$; $A_Q(t)\delta(t)$	$\widetilde{\mathfrak{g}}_g[u]$; $A_g(\sigma_1,\ldots,\sigma_{2g})\delta(t)$
Coho. Hall algebra $H^T_*(T^*\mathfrak{M})$	$\widetilde{\mathfrak{g}}_g[u^{\pm 1}]$; $A_Q(t)/(1-t)$	$\widetilde{\mathfrak{g}}_g[u]$; $A_g(\sigma_1,\ldots,\sigma_{2g})/(1-t)$
Hall algebra **H**	$\widetilde{\mathfrak{g}}_Q$; $A_Q(t)$	$\widetilde{\mathfrak{g}}_g$; $A_g(\sigma_1,\ldots,\sigma_{2g})$
Spherical Hall algebra $\mathbf{H}^{\text{sph}}$	$\mathfrak{g}_Q$; $A_Q(0)$	$\mathfrak{g}_g$; $A_g(0,\ldots,0)$

Acknowledgments. It is, of course, impossible to thank here all the people whose work had a strong influence on the research presented here. I would nevertheless like to express my gratitude to the following people for their help throughout the years: R. Bezrukavnikov,

T. Bozec, P.-H. Chaudouard, I. Cherednik, W. Crawley-Boevey, B. Davison, P. Etingof, B. Feigin, D. Fratila, I. Frenkel, T. Hausel, V. Kac, S.-J. Kang, M. Kapranov, M. Kashiwara, M. Khovanov, G. Laumon, E. Letellier, G. Lusztig, S. Meinhardt, A. Mellit, A. Minets, S. Mozgovoy, H. Nakajima, A. Negut, A. Okounkov, C. Ringel, F. Rodriguez-Villegas, F. Sala, P. Samuelson, Y. Soibelman, M. Reineke, E. Vasserot, and B. Webster.

References

Pierre Baumann and Christian Kassel (2001). "The Hall algebra of the category of coherent sheaves on the projective line". *J. Reine Angew. Math.* 533, pp. 207–233. MR: 1823869 (cit. on p. 1432).

Francois Bergeron, Adriano Garsia, Emily Sergel Leven, and Guoce Xin (2016). "Some remarkable new plethystic operators in the theory of Macdonald polynomials". *J. Comb.* 7.4, pp. 671–714. MR: 3538159 (cit. on p. 1432).

T. Bozec and O. Schiffmann (Oct. 2017). "Counting cuspidals for quivers". arXiv: 1710.03036 (cit. on pp. 1421, 1422).

T. Bozec, O. Schiffmann, and E. Vasserot (Jan. 2017). "On the number of points of nilpotent quiver varieties over finite fields". arXiv: 1701.01797 (cit. on pp. 1419, 1424, 1429).

Tristan Bozec (2015). "Quivers with loops and perverse sheaves". *Math. Ann.* 362.3-4, pp. 773–797. MR: 3368082 (cit. on p. 1419).

– (2016). "Quivers with loops and generalized crystals". *Compos. Math.* 152.10, pp. 1999–2040. MR: 3569998 (cit. on p. 1419).

– (Dec. 2017). "Irreducible components of the global nilpotent cone". arXiv: 1712.07362 (cit. on p. 1436).

Igor Burban and Olivier Schiffmann (2012). "On the Hall algebra of an elliptic curve, I". *Duke Math. J.* 161.7, pp. 1171–1231. MR: 2922373 (cit. on p. 1432).

Sabin Cautis, Aaron D. Lauda, Anthony Licata, Peter Samuelson, and Joshua Sussan (Sept. 2016). "The Elliptic Hall algebra and the deformed Khovanov Heisenberg category". arXiv: 1609.03506 (cit. on p. 1433).

William Crawley-Boevey and Michel Van den Bergh (2004). "Absolutely indecomposable representations and Kac-Moody Lie algebras". *Invent. Math.* 155.3. With an appendix by Hiraku Nakajima, pp. 537–559. MR: 2038196 (cit. on pp. 1412, 1416).

Ben Davison (Feb. 2016). "The integrality conjecture and the cohomology of preprojective stacks". arXiv: 1602.02110 (cit. on pp. 1424, 1428).

Ben Davison and Sven Meinhardt (Jan. 2016). "Cohomological Donaldson-Thomas theory of a quiver with potential and quantum enveloping algebras". arXiv: 1601.02479 (cit. on p. 1428).

Pierre Deligne (2015). "Comptage de faisceaux l-adiques". *Astérisque* 369, pp. 285–312. MR: 3379637 (cit. on p. 1435).

Bangming Deng and Jie Xiao (2003). "A new approach to Kac's theorem on representations of valued quivers". *Math. Z.* 245.1, pp. 183–199. MR: 2023959 (cit. on p. 1421).

Jintai Ding and Kenji Iohara (1997). "Generalization of Drinfeld quantum affine algebras". *Lett. Math. Phys.* 41.2, pp. 181–193. MR: 1463869 (cit. on p. 1433).

Galyna Dobrovolska, Victor Ginzburg, and Roman Travkin (Dec. 2016). "Moduli spaces, indecomposable objects and potentials over a finite field". arXiv: 1612.01733 (cit. on p. 1437).

V. G. Drinfel'd (1981). "The number of two-dimensional irreducible representations of the fundamental group of a curve over a finite field". *Funktsional. Anal. i Prilozhen.* 15.4, pp. 75–76. MR: 639205 (cit. on p. 1435).

Roman Fedorov, Alexander Soibelman, and Yan Soibelman (May 2017). "Motivic classes of moduli of Higgs bundles and moduli of bundles with connections". arXiv: 1705.04890 (cit. on p. 1437).

B. Feigin, E. Feigin, M. Jimbo, T. Miwa, and E. Mukhin (2011). "Quantum continuous $\mathfrak{gl}_\infty$: semiinfinite construction of representations". *Kyoto J. Math.* 51.2, pp. 337–364. MR: 2793271 (cit. on p. 1433).

Philippe Di Francesco and Rinat Kedem (Apr. 2017). "(t,q) Q-systems, DAHA and quantum toroidal algebras via generalized Macdonald operators". arXiv: 1704.00154 (cit. on p. 1432).

Dragos Fratila (2013). "Cusp eigenforms and the Hall algebra of an elliptic curve". *Compos. Math.* 149.6, pp. 914–958. MR: 3077657 (cit. on p. 1435).

Peter Gabriel (1972). "Unzerlegbare Darstellungen. I". *Manuscripta Math.* 6, 71–103, correction, ibid. 6 (1972), 309. MR: 0332887 (cit. on p. 1416).

Eugene Gorsky and Andrei Neguț (2015). "Refined knot invariants and Hilbert schemes". *J. Math. Pures Appl. (9)* 104.3, pp. 403–435. MR: 3383172 (cit. on p. 1432).

James A. Green (1995). "Hall algebras, hereditary algebras and quantum groups". *Invent. Math.* 120.2, pp. 361–377. MR: 1329046 (cit. on pp. 1417, 1418).

Tamás Hausel (2010). "Kac's conjecture from Nakajima quiver varieties". *Invent. Math.* 181.1, pp. 21–37. MR: 2651380 (cit. on pp. 1416, 1429).

– (2013). "Global topology of the Hitchin system". In: *Handbook of moduli. Vol. II.* Vol. 25. Adv. Lect. Math. (ALM). Int. Press, Somerville, MA, pp. 29–69. MR: 3184173 (cit. on p. 1437).

Tamás Hausel, Emmanuel Letellier, and Fernando Rodriguez-Villegas (2013). "Positivity for Kac polynomials and DT-invariants of quivers". *Ann. of Math. (2)* 177.3, pp. 1147–1168. MR: 3034296 (cit. on p. 1412).

Tamás Hausel and Fernando Rodriguez-Villegas (2008). "Mixed Hodge polynomials of character varieties". *Invent. Math.* 174.3. With an appendix by Nicholas M. Katz, pp. 555–624. MR: 2453601 (cit. on pp. 1415, 1437).

Jochen Heinloth (2012). "Cohomology of the moduli stack of coherent sheaves on a curve". In: *Geometry and arithmetic*. EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, pp. 165–171. MR: 2987660 (cit. on p. 1436).

Jiuzhao Hua (2000). "Counting representations of quivers over finite fields". *J. Algebra* 226.2, pp. 1011–1033. MR: 1752774 (cit. on pp. 1412, 1414).

V. G. Kac (1980). "Infinite root systems, representations of graphs and invariant theory". *Invent. Math.* 56.1, pp. 57–92. MR: 557581 (cit. on p. 1411).

Victor G. Kac (1985). *Infinite-dimensional Lie algebras*. Second. Cambridge University Press, Cambridge, pp. xviii+280. MR: 823672 (cit. on p. 1416).

Joel Kamnitzer (2014). "Categorification of Lie algebras (after Rouquier, Khovanov-Lauda, . . .)". *Astérisque* 361, Exp. No. 1072, x, 397–419. MR: 3289289.

Seok-Jin Kang (2018). "Ringel–Hall algebra construction of quantum Borcherds–Bozec algebras". *J. Algebra* 498, pp. 344–361. arXiv: 1709.10235. MR: 3754419 (cit. on p. 1419).

Seok-Jin Kang, Masaki Kashiwara, and Olivier Schiffmann (2009). "Geometric construction of crystal bases for quantum generalized Kac-Moody algebras". *Adv. Math.* 222.3, pp. 996–1015. MR: 2553376 (cit. on p. 1419).

Seok-Jin Kang and Olivier Schiffmann (2006). "Canonical bases for quantum generalized Kac-Moody algebras". *Adv. Math.* 200.2, pp. 455–478. MR: 2200853 (cit. on p. 1419).

M. M. Kapranov (1997). "Eisenstein series and quantum affine algebras". *J. Math. Sci. (New York)* 84.5. Algebraic geometry, 7, pp. 1311–1360. MR: 1465518 (cit. on p. 1432).

M. Kapranov, O. Schiffmann, and E. Vasserot (2014). "Spherical Hall algebra of $\overline{\mathrm{Spec}(\mathbb{Z})}$". In: *Homological mirror symmetry and tropical geometry*. Vol. 15. Lect. Notes Unione Mat. Ital. Springer, Cham, pp. 153–196. MR: 3330787 (cit. on p. 1434).

Mikhail Kapranov, Olivier Schiffmann, and Eric Vasserot (2017). "The Hall algebra of a curve". *Selecta Math. (N.S.)* 23.1, pp. 117–177. MR: 3595890 (cit. on p. 1434).

Masaki Kashiwara and Yoshihisa Saito (1997). "Geometric construction of crystal bases". *Duke Math. J.* 89.1, pp. 9–36. MR: 1458969 (cit. on p. 1418).

Mikhail Khovanov and Aaron D. Lauda (2009). "A diagrammatic approach to categorification of quantum groups. I". *Represent. Theory* 13, pp. 309–347. MR: 2525917 (cit. on p. 1423).

Maxim Kontsevich (2009). "Notes on motives in finite characteristic". In: *Algebra, arithmetic, and geometry: in honor of Yu. I. Manin. Vol. II*. Vol. 270. Progr. Math. Birkhäuser Boston, Inc., Boston, MA, pp. 213–247. MR: 2641191 (cit. on p. 1435).

Maxim Kontsevich and Yan Soibelman (2011). "Cohomological Hall algebra, exponential Hodge structures and motivic Donaldson-Thomas invariants". *Commun. Number Theory Phys.* 5.2, pp. 231–352. MR: 2851153 (cit. on p. 1425).

Laurent Lafforgue (2002). "Chtoucas de Drinfeld et correspondance de Langlands". *Invent. Math.* 147.1, pp. 1–241. MR: 1875184 (cit. on p. 1435).

Jyun-Ao Lin (Oct. 2014). "Spherical Hall algebras of a weighted projective curve". arXiv: 1410.0896 (cit. on pp. 1415, 1434).

G. Lusztig (1991). "Quivers, perverse sheaves, and quantized enveloping algebras". *J. Amer. Math. Soc.* 4.2, pp. 365–421. MR: 1088333 (cit. on p. 1418).

– (1992). "Affine quivers and canonical bases". *Inst. Hautes Études Sci. Publ. Math.* 76, pp. 111–163. MR: 1215594.

George Lusztig (2010). *Introduction to quantum groups*. Modern Birkhäuser Classics. Reprint of the 1994 edition. Birkhäuser/Springer, New York, pp. xiv+346. MR: 2759715.

Davesh Maulik and Andrei Okounkov (Nov. 2012). "Quantum Groups and Quantum Cohomology". arXiv: 1211.1287 (cit. on p. 1430).

Anton Mellit (Oct. 2017a). "Poincare polynomials of character varieties, Macdonald polynomials and affine Springer fibers". arXiv: 1710.04513 (cit. on p. 1415).

– (July 2017b). "Poincaré polynomials of moduli spaces of Higgs bundles and character varieties (no punctures)". arXiv: 1707.04214 (cit. on p. 1414).

Kei Miki (2007). "A (q, γ) analog of the $W_{1+\infty}$ algebra". *J. Math. Phys.* 48.12, pp. 123520, 35. MR: 2377852 (cit. on p. 1433).

Aliaksandr Minets (Jan. 2018). "Cohomological Hall algebras for Higgs torsion sheaves, moduli of triples and Hilbert schemes on surfaces". arXiv: 1801.01429 (cit. on pp. 1436, 1437).

Hugh Morton and Peter Samuelson (2017). "The HOMFLYPT skein algebra of the torus and the elliptic Hall algebra". *Duke Math. J.* 166.5, pp. 801–854. MR: 3626565 (cit. on p. 1432).

Sergey Mozgovoy (July 2011). "Motivic Donaldson-Thomas invariants and McKay correspondence". arXiv: 1107.6044 (cit. on pp. 1421, 1424).

Sergey Mozgovoy and Olivier Schiffmann (May 2017). "Counting Higgs bundles and type A quiver bundles". arXiv: 1705.04849 (cit. on pp. 1437, 1438).

Hiraku Nakajima (1994). "Instantons on ALE spaces, quiver varieties, and Kac-Moody algebras". *Duke Math. J.* 76.2, pp. 365–416. MR: 1302318 (cit. on p. 1424).

– (1998). "Quiver varieties and Kac-Moody algebras". *Duke Math. J.* 91.3, pp. 515–560. MR: 1604167 (cit. on p. 1429).

– (2001). "Quiver varieties and finite-dimensional representations of quantum affine algebras". *J. Amer. Math. Soc.* 14.1, pp. 145–238. MR: 1808477 (cit. on p. 1429).

Andrei Neguţ (2015). "Moduli of flags of sheaves and their K-theory". *Algebr. Geom.* 2.1, pp. 19–43. MR: 3322196 (cit. on p. 1427).

– (Aug. 2016). "The q-AGT-W relations via shuffle algebras". arXiv: 1608.08613 (cit. on pp. 1427, 1432).

– (Nov. 2017). "AGT relations for sheaves on surfaces". arXiv: 1711.00390 (cit. on p. 1437).

A Okounkov (n.d.). *On some interesting Lie algebras. Talk at the Conference in Honor of V. Kac, IMPA,(2013)* (cit. on p. 1430).

Jie Ren and Yan Soibelman (2017). "Cohomological Hall algebras, semicanonical bases and Donaldson-Thomas invariants for 2-dimensional Calabi-Yau categories (with an appendix by Ben Davison)". In: *Algebra, geometry, and physics in the 21st century*. Vol. 324. Progr. Math. Birkhäuser/Springer, Cham, pp. 261–293. MR: 3727563 (cit. on p. 1425).

Claus Michael Ringel (1990). "Hall algebras and quantum groups". *Invent. Math.* 101.3, pp. 583–591. MR: 1062796 (cit. on pp. 1417, 1418).

Raphaël Rouquier (2012). "Quiver Hecke algebras and 2-Lie algebras". *Algebra Colloq.* 19.2, pp. 359–410. MR: 2908731 (cit. on p. 1423).

Francesco Sala and Olivier Schiffmann (Jan. 2018). "Cohomological Hall algebra of Higgs sheaves on a curve". arXiv: 1801.03482 (cit. on p. 1436).

O. Schiffmann and E. Vasserot (2011). "The elliptic Hall algebra, Cherednik Hecke algebras and Macdonald polynomials". *Compos. Math.* 147.1, pp. 188–234. MR: 2771130 (cit. on p. 1432).

– (2012). "Hall algebras of curves, commuting varieties and Langlands duality". *Math. Ann.* 353.4, pp. 1399–1451. MR: 2944034 (cit. on pp. 1433, 1434).

– (2013a). "Cherednik algebras, W-algebras and the equivariant cohomology of the moduli space of instantons on $\mathbf{A}^2$". *Publ. Math. Inst. Hautes Études Sci.* 118, pp. 213–342. MR: 3150250 (cit. on pp. 1425, 1432, 1439).

Olivier Schiffmann (2004). "Noncommutative projective curves and quantum loop algebras". *Duke Math. J.* 121.1, pp. 113–168. MR: 2031167 (cit. on pp. 1415, 1434).

– (2008). "Variétés carquois de Nakajima (d'après Nakajima, Lusztig, Varagnolo, Vasserot, Crawley-Boevey, et al.)" *Astérisque* 317. Séminaire Bourbaki. Vol. 2006/2007, Exp. No. 976, ix, 295–344. MR: 2487738 (cit. on p. 1429).

– (2011). "Spherical Hall algebras of curves and Harder-Narasimhan stratas". *J. Korean Math. Soc.* 48.5, pp. 953–967. MR: 2850072 (cit. on p. 1432).

– (2012). "Lectures on Hall algebras". In: *Geometric methods in representation theory. II*. Vol. 24. Sémin. Congr. Soc. Math. France, Paris, pp. 1–141. MR: 3202707 (cit. on p. 1418).

– (2016). "Indecomposable vector bundles and stable Higgs bundles over smooth projective curves". *Ann. of Math. (2)* 183.1, pp. 297–362. MR: 3432585 (cit. on pp. 1413, 1414, 1437, 1438).

Olivier Schiffmann and Eric Vasserot (2013b). "The elliptic Hall algebra and the K-theory of the Hilbert scheme of $\mathbb{A}^2$". *Duke Math. J.* 162.2, pp. 279–366. MR: 3018956 (cit. on p. 1432).

– (May 2017a). "On cohomological Hall algebras of quivers : generators". arXiv: 1705.07488 (cit. on pp. 1424, 1425, 1427, 1429).

– (May 2017b). "On cohomological Hall algebras of quivers : Yangians". arXiv: 1705.07491 (cit. on p. 1430).

Bert Sevenhant and Michel Van Den Bergh (2001). "A relation between a conjecture of Kac and the structure of the Hall algebra". *J. Pure Appl. Algebra* 160.2-3, pp. 319–332. MR: 1836006 (cit. on p. 1420).

Michela Varagnolo (2000). "Quiver varieties and Yangians". *Lett. Math. Phys.* 53.4, pp. 273–283. MR: 1818101 (cit. on p. 1429).

Yaping Yang and Gufang Zhao (July 2014). "The cohomological Hall algebra of a preprojective algebra". arXiv: 1407.7994 (cit. on pp. 1425, 1430).

– (Apr. 2016). "On two cohomological Hall algebras". arXiv: 1604.01477 (cit. on p. 1425).

H. Yu (n.d.). In preparation (cit. on p. 1435).

Received 2018-02-26.

Olivier G. Schiffmann
olivier.schiffmann@gmail.com

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1443–1464)

MODULI SPACES OF LOCAL G-SHTUKAS

Eva Viehmann

Abstract

We give an overview of the theory of local **G**-shtukas and their moduli spaces that were introduced in joint work of U. Hartl and the author, and in the past years studied by many people. We also discuss relations to moduli of global **G**-shtukas, properties of their special fiber through affine Deligne-Lusztig varieties and of their generic fiber, such as the period map.

1 Introduction

Local **G**-shtukas are an analog over local function fields of p-divisible groups with additional structure. In this article we give an overview about the theory of moduli spaces of local **G**-shtukas and their relation to moduli of global $\mathfrak{G}$-shtukas. It parallels the theory of Rapoport-Zink moduli spaces of p-divisible groups and their relation to Shimura varieties. Yet it has the additional charm that additional structure given by any parahoric group scheme **G** can be easily encoded and treated in a group-theoretic way.

We begin by defining local **G**-shtukas for a parahoric group scheme **G** over Spec $\mathbb{F}_q[\![z]\!]$ (see Section 2). They are pairs consisting of an $L^+\mathbf{G}$-torsor $\mathcal{G}$ together with an isomorphism $\sigma^*L\mathcal{G} \to L\mathcal{G}$ for the associated $L\mathbf{G}$-torsor. Here for unexplained notation we refer to the respective sections. Local **G**-shtukas are the function field analog of p-divisible groups with additional structure, as well as a group theoretic generalization of Drinfeld's shtukas. Next we discuss possible bounds on the singularities of local **G**-shtukas. These are a more general replacement of the minuscule coweights fixed to define Shimura varieties. In Section 3 we consider deformations for local **G**-shtukas, and the description of the universal deformation in terms of the loop group $L\mathbf{G}$. In Section 4 we describe the analog of Rapoport-Zink moduli spaces for local **G**-shtukas bounded by a given bound $\hat{Z}$ and within a given quasi-isogeny class. They are representable by a formal scheme, locally formally of finite type over Spf$\breve{R}$. Here, $\breve{R}$ is the completion of the maximal unramified extension of the ring of definition of the chosen bound $\hat{Z}$. Also,

MSC2010: primary 11G09; secondary 11G18, 14L05, 14M15.

they admit a tower of coverings indexed by compact open subgroups of $\mathbf{G}(\mathbb{F}_q(\!(z)\!))$. Our next topic is the special fiber of a Rapoport-Zink moduli space of local **G**-shtukas. It can be identified with a so-called affine Deligne-Lusztig variety. In Section 5 we discuss applications of this result to geometric questions such as dimensions and closure relations of Newton strata. In Section 6 we describe the relation between our local **G**-shtukas and global $\mathfrak{G}$-shtukas where $\mathfrak{G}$ is a parahoric group scheme over a smooth projective geometrically irreducible curve C over $\mathbb{F}_q$. There are several results in this direction, the central idea being that one wants to describe a global $\mathfrak{G}$-shtuka by corresponding local $\mathbf{G}_i$-shtukas associated with each of the legs c_i. Due to our choice of definition of local **G**-shtukas, this works particularly well when considering global $\mathfrak{G}$-shtukas with fixed legs. Then one obtains analogues/generalizations of Serre-Tate's theorem, as well as of several classical results comparing Shimura varieties and Rapoport-Zink spaces in the arithmetic case. In the last section we define period spaces as suitable subspaces of an affine Grassmannian. Contrary to the arithmetic case they are no longer subspaces of a classical flag variety because we allow also non-minuscule bounds. We define the period map from the analytic space associated with a Rapoport-Zink space to the corresponding period space, and discuss its image and compatibility with the tower of coverings of the generic fiber of the Rapoport-Zink space.

2 Local G-shtukas

2.1 Generalities. Let $\mathbb{F}_q$ be a finite field of characteristic p with q elements, let $\mathbb{F}$ be a fixed algebraic closure of $\mathbb{F}_q$, and let $\mathbb{F}_q[\![z]\!]$ and $\mathbb{F}_q[\![\zeta]\!]$ be the power series rings over $\mathbb{F}_q$ in the (independent) variables z and ζ. As base schemes we will consider the category $\mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}$ consisting of schemes over $\operatorname{Spec}\mathbb{F}_q[\![\zeta]\!]$ on which ζ is locally nilpotent. Let **G** be a parahoric group scheme over $\operatorname{Spec}\mathbb{F}_q[\![z]\!]$ with connected reductive generic fiber, compare Bruhat and Tits [1972], Déf. 5.2.6 and Haines and Rapoport [2008].

Let $S \in \mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}$ and consider any sheaf of groups H on S for the fpqc-topology. By an H-torsor on S we mean a sheaf $\mathcal{H}$ for the fpqc-topology on S together with a (right) action of the sheaf H such that $\mathcal{H}$ is isomorphic to H on an fpqc-covering of S.

Let now $L\mathbf{G}$ and $L^+\mathbf{G}$ be the loop group and the group of positive loops associated with **G**, i.e. for an $\mathbb{F}_q$-algebra R let

$$(L^+\mathbf{G})(R) = \mathbf{G}(R[\![z]\!]) \quad \text{and} \quad (L\mathbf{G})(R) = \mathbf{G}(R(\!(z)\!)).$$

Let $\mathcal{G}$ be an $L^+\mathbf{G}$-torsor on S. Via the inclusion $L^+\mathbf{G} \subset L\mathbf{G}$ we can associate an $L\mathbf{G}$-torsor $L\mathcal{G}$ with $\mathcal{G}$. For any $L\mathbf{G}$-torsor $\mathcal{G}'$ on S we denote by $\sigma^*\mathcal{G}'$ the pullback of $\mathcal{G}'$ under the q-Frobenius morphism $\sigma := \mathrm{Frob}_q : S \to S$.

Definition 2-1. A *local* **G**-*shtuka* over some $S \in \mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}$ is a pair $\underline{\mathcal{G}} = (\mathcal{G}, \tau_{\mathcal{G}})$ consisting of an $L^+\mathbf{G}$-torsor $\mathcal{G}$ on S and an isomorphism of the associated $L\mathbf{G}$-torsors $\tau_{\mathcal{G}} \colon \sigma^* L\mathcal{G} \xrightarrow{\sim} L\mathcal{G}$.

A *quasi-isogeny* $g \colon (\mathcal{G}', \tau_{\mathcal{G}'}) \to (\mathcal{G}, \tau_{\mathcal{G}})$ between local **G**-shtukas over S is an isomorphism $g \colon L\mathcal{G}' \xrightarrow{\sim} L\mathcal{G}$ of the associated $L\mathbf{G}$-torsors with $g \circ \tau_{\mathcal{G}'} = \tau_{\mathcal{G}} \circ \sigma^* g$.

Local **G**-shtukas were introduced and studied in Hartl and Viehmann [2011], Hartl and Viehmann [2012] in the case where **G** is a constant split reductive group over $\mathbb{F}_q$. The general case was first considered in work of Arasteh Rad and Hartl [2014].

Example 2-2. For $\mathbf{G} = \mathrm{GL}_r$, we have the following more classical description.

A local shtuka over $S \in \mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}$ (of rank r) is a pair (M, ϕ) where M is a sheaf of $\mathcal{O}_S[\![z]\!]$-modules on S which Zariski-locally is free of rank r, together with an isomorphism of $\mathcal{O}_S(\!(z)\!)$-modules

$$\phi : \sigma^* M \otimes_{\mathcal{O}_S[\![z]\!]} \mathcal{O}_S(\!(z)\!) \xrightarrow{\sim} M \otimes_{\mathcal{O}_S[\![z]\!]} \mathcal{O}_S(\!(z)\!).$$

Then the category of local GL_r-shtukas over S is equivalent to the category of local shtukas of rank r over S, see Hartl and Viehmann [2011], Lemma 4.2.

Local shtukas were first introduced by Anderson [1993] over a complete discrete valuation ring. Genestier [1996] constructed moduli spaces for them in the Drinfeld case and used these to uniformize Drinfeld modular varieties.

An important invariant of local **G**-shtukas is their *Newton point*. To define it let k be an algebraically closed field of characteristic p and let $L = k(\!(z)\!)$. Then every local **G**-shtuka $(\mathcal{G}, \tau_{\mathcal{G}})$ over k has a trivialization $\mathcal{G} \cong (L^+\mathbf{G})_k$. Via this isomorphism the Frobenius map $\tau_{\mathcal{G}}$ corresponds to an element $b \in L\mathbf{G}(k) = \mathbf{G}(L)$. Changing the trivialization replaces b by a different representative of its $L^+\mathbf{G}(k)$-σ-conjugacy class. In the same way, local **G**-shtukas isogenous to $(\mathcal{G}, \tau_{\mathcal{G}})$ correspond to the elements of the σ-conjugacy class

$$[b] = \{g^{-1} b \sigma(g) \mid g \in \mathbf{G}(L)\}.$$

Let G denote the generic fiber of **G**. Then the set of σ-conjugacy classes $B(G) = \{[b] \mid b \in G(L)\}$ for quasi-split G is described by Kottwitz [1985], Kottwitz [1997] by two invariants. Let T be a maximal torus of G. The first invariant is then the Kottwitz point, i.e. the image under the Kottwitz map $\kappa_G : G(L) \to \pi_1(G)_\Gamma$. Here, $\pi_1(G)$ is Borovoi's fundamental group, and Γ is the absolute Galois group of $\mathbb{F}_q(\!(z)\!)$. This invariant is locally constant on LG, and can also be computed using the identification $\pi_0(LG) = \pi_1(G)/\Gamma$, compare Neupert [2016], 2.2. The second invariant is the Newton point ν_b, an element of $X_*(T)^\Gamma_{\mathbb{Q}}$.

Consider now a local $\mathbf{G}$-shtuka $(\mathcal{G}, \tau_{\mathcal{G}})$ over a scheme S. Then we obtain an induced decomposition of $S(k)$ into subsets $\mathfrak{N}_{[b]}(k)$ for $[b] \in B(G)$ with

$$\mathfrak{N}_{[b]}(k) = \{x \in S(k) \mid (\mathcal{G}, \tau_{\mathcal{G}})_x \text{ is in the isogeny class of } [b]\}.$$

By Rapoport and Richartz [1996], this induces a decomposition of S into locally closed subschemes, which we call Newton strata. If S is connected, they correspond to the strata obtained by fixing the Newton point only.

2.2 Bounds. To ensure finiteness properties of moduli spaces of local $\mathbf{G}$-shtukas, or of deformations of local $\mathbf{G}$-shtukas, we bound the singularity of the morphism $\tau_{\mathcal{G}}$. In the arithmetic context, such bounds are usually assumed to be minuscule and correspond to the choice of the Hodge cocharacter defining a Shimura datum. In our context we define bounds more generally as suitable ind-subschemes of the affine flag variety.

The *affine flag variety* $\mathrm{Flag}_{\mathbf{G}}$ *associated with* G is the fpqc-sheaf associated with the presheaf

$$S \mapsto L\mathbf{G}(S)/L^+\mathbf{G}(S) = \mathbf{G}(\mathcal{O}_S(\!(z)\!)(S))/\mathbf{G}(\mathcal{O}_S[\![z]\!](S))$$

on the category of $\mathbb{F}_q$-schemes. By Pappas and Rapoport [2008], Theorem 1.4 and Richarz [2016] Theorem A, $\mathrm{Flag}_{\mathbf{G}}$ is represented by an ind-scheme which is ind-projective over $\mathbb{F}_q$.

Consider further the group scheme $\mathbf{G} \times_{\mathbb{F}_q[\![z]\!]} \mathrm{Spec}\, \mathbb{F}_q(\!(\zeta)\!)[\![z-\zeta]\!]$ under the homomorphism

$$\mathbb{F}_q[\![z]\!] \to \mathbb{F}_q(\!(\zeta)\!)[\![z-\zeta]\!], \qquad z \mapsto z = \zeta + (z-\zeta). \tag{2-3}$$

The associated affine Grassmannian $\mathrm{Gr}_{\mathbf{G}}^{B_{\mathrm{dR}}}$ is the sheaf of sets for the fpqc-topology on $\mathrm{Spec}\, \mathbb{F}_q(\!(\zeta)\!)$ associated with the presheaf

$$X \mapsto \mathbf{G}(\mathcal{O}_X(\!(z-\zeta)\!))/\mathbf{G}(\mathcal{O}_X[\![z-\zeta]\!]). \tag{2-4}$$

Remark 2-5. $\mathrm{Gr}_{\mathbf{G}}^{B_{\mathrm{dR}}}$ is in the same way as $\mathrm{Flag}_{\mathbf{G}}$ above an ind-projective ind-scheme over $\mathrm{Spec}\, \mathbb{F}_q(\!(\zeta)\!)$. Here, note that the homomorphism in (2-3) induces an inclusion $\mathbb{F}_q(\!(z)\!) \to \mathbb{F}_q(\!(\zeta)\!)[\![z-\zeta]\!]$, so the group $\mathbf{G} \times_{\mathbb{F}_q[\![z]\!]} \mathrm{Spec}\, \mathbb{F}_q(\!(\zeta)\!)[\![z-\zeta]\!]$ is reductive, justifying the name affine "Grassmannian". The notation B_{dR} refers to the fact that if C is the completion of an algebraic closure of $\mathbb{F}_q(\!(\zeta)\!)$, then $C(\!(z-\zeta)\!)$ is the function field analog of Fontaine's p-adic period field B_{dR}.

We fix an algebraic closure $\overline{\mathbb{F}_q(\!(\zeta)\!)}$ of $\mathbb{F}_q(\!(\zeta)\!)$, and consider pairs $(R, \hat{Z}_R)$, where $R/\mathbb{F}_q[\![\zeta]\!]$ is a finite extension of discrete valuation rings contained in $\overline{\mathbb{F}_q(\!(\zeta)\!)}$, and where $\hat{Z}_R \subset \widehat{\mathrm{Flag}}_{\mathbf{G},R} := \mathrm{Flag}_{\mathbf{G}} \widehat{\times}_{\mathbb{F}_q} \mathrm{Spf} R$ is a closed ind-subscheme that is contained in a bounded (or: projective) subscheme of $\widehat{\mathrm{Flag}}_{\mathbf{G},R}$. Two such pairs $(R, \hat{Z}_R)$ and $(R', \hat{Z}'_{R'})$ are equivalent if they agree over a finite extension of the rings R, R'.

A *bound* is then abstractly defined as an equivalence class $\hat{Z} := [(R, \hat{Z}_R)]$ of such pairs $(R, \hat{Z}_R)$ satisfying certain properties (for the precise conditions compare Hartl and Viehmann [2017], Def. 2.1), in particular

1. $\hat{Z}_R \subset \widehat{\mathrm{Flag}}_{\mathbf{G},R}$ is a ζ-adic formal scheme over $\mathrm{Spf} R$ that is stable under the left $L^+\mathbf{G}$-action.
2. The special fiber $Z_R := \hat{Z}_R \widehat{\times}_{\mathrm{Spf} R} \mathrm{Spec}\, \kappa_R$ is a quasi-compact subscheme of $\mathrm{Flag}_G \widehat{\times}_{\mathbb{F}_q} \kappa_R$ where κ_R is the residue field of R.
3. Let $\hat{Z}_R^{\mathrm{an}}$ be the strictly $R[\frac{1}{\zeta}]$-analytic space associated with $\hat{Z}_R$. Then the $\hat{Z}_R^{\mathrm{an}}$ are invariant under the left multiplication of $\mathbf{G}(\cdot[\![z-\zeta]\!])$ on $\mathrm{Gr}_{\mathbf{G}}^{\mathbf{B}_{\mathrm{dR}}}$.

The reflex ring $R_{\hat{Z}}$ of $[(R, \hat{Z}_R)]$ is the intersection in $\overline{\mathbb{F}_q(\!(\zeta)\!)}$ of the fixed field of $\{\gamma \in \mathrm{Aut}_{\mathbb{F}_q[\![\zeta]\!]}(\overline{\mathbb{F}_q(\!(\zeta)\!)}) \mid \gamma(\hat{Z}) = \hat{Z}\}$ with all finite extensions of $\mathbb{F}_q[\![\zeta]\!]$ over which a representative $\hat{Z}_R$ of the given bound exists.

Let $(\mathcal{G}, \tau_{\mathcal{G}})$ be a local **G**-shtuka over some $S \in \mathrm{Nilp}_{R_{\hat{Z}}}$ and let S' be an étale covering of S over which a trivialization $\mathcal{G} \cong (L^+\mathbf{G})_{S'}$ exists. Then $(\mathcal{G}, \tau_{\mathcal{G}})$ (or $\tau_{\mathcal{G}}$) is *bounded by* $\hat{Z}$ if for every such trivialization and for every finite extension R of $\mathbb{F}_q[\![\zeta]\!]$ over which a representative $\hat{Z}_R$ of $\hat{Z}$ exists, the morphism

$$S' \widehat{\times}_{R_{\hat{Z}}} \mathrm{Spf} R \to L\mathbf{G} \widehat{\times}_{\mathbb{F}_q} \mathrm{Spf} R \to \widehat{\mathrm{Flag}}_{\mathbf{G},R}$$

induced by $\tau_{\mathcal{G}}$ factors through $\hat{Z}_R$.

The above definition of bounds follows the strategy to allow as general bounds as possible. On the other hand, classically and by analogy with the arithmetic case, one considers bounds given as Schubert varieties. They are described as follows (compare Arasteh Rad and Hartl [2014], Example 4.12).

Example 2-6. Consider the base change G_L of G to $L = \mathbb{F}(\!(z)\!)$. Let A be a maximal split torus in G_L and let T be its centralizer, a maximal torus. Let $N = N(T)$ be the normalizer of T and let $\mathcal{T}^0$ be the identity component of the Néron model of T over $\mathcal{O}_L = \mathbb{F}[\![z]\!]$.

Consider the Iwahori–Weyl group $\widetilde{W} = N(L)/\mathcal{T}^0(\mathcal{O}_L)$ and let $\widetilde{W}^{\mathbf{G}} = (N(L) \cap \mathbf{G}(\mathcal{O}_L))/\mathcal{T}^0(\mathcal{O}_L)$. Then by the Bruhat-Tits decomposition we have a bijection

$$\widetilde{W}^{\mathbf{G}} \backslash \widetilde{W} / \widetilde{W}^{\mathbf{G}} \to L^+\mathbf{G}(\mathbb{F}) \backslash L\mathbf{G}(\mathbb{F}) / L^+\mathbf{G}(\mathbb{F}). \tag{2-7}$$

Let ω be in the right hand side, and let $\mathbb{F}_\omega$ be a finite extension of $\mathbb{F}_q$ such that ω has a representative $g_\omega \in L\mathbf{G}(\mathbb{F}_\omega)$. We define the *Schubert variety* S_ω as the ind-scheme theoretic closure of the $L^+\mathbf{G}$-orbit of g_ω in $\mathrm{Flag}_{\mathbf{G}} \widehat{\times}_{\mathbb{F}_q} \mathbb{F}_\omega$. Let $R = \mathbb{F}_\omega[\![\zeta]\!]$ and let $\hat{Z}_{\mathbb{F}_\omega[\![\zeta]\!]} =$

$S_\omega \widehat{\times}_{\mathbb{F}_\omega} \mathrm{Spf} R$. Then the equivalence class of $(R, \hat{Z}_R)$ defines a bound. In this case we also say "bounded by ω" instead of "bounded by $(R, \hat{Z}_R)$".

3 Deformations

Let $\mathrm{Flag}_{\mathbf{G}}$ denote again the affine flag variety, i.e. the quotient sheaf $\mathrm{Flag}_{\mathbf{G}} = L\mathbf{G}/L^+\mathbf{G}$, an ind-scheme over $\mathbb{F}_q$ which is of ind-finite type. Let $\widehat{\mathrm{Flag}}_{\mathbf{G}}$ be the fiber product $\mathrm{Flag}_{\mathbf{G}} \times_{\mathrm{Spec}\ \mathbb{F}_q} \mathrm{Spf}\ \mathbb{F}_q(\!(\zeta)\!)$.

Then generalizing Hartl and Viehmann [2011, Prop. 5.1] (where the case that $\mathbf{G}$ is split is considered) the ind-scheme $\widehat{\mathrm{Flag}}_{\mathbf{G}}$ pro-represents the functor $\mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}{}^o \to (\mathrm{Sets})$

$$\begin{aligned} S \mapsto & \{(\mathcal{G}, \delta) \mid \mathcal{G} \text{ a } \mathbf{G}\text{-torsor on } S, \\ & \quad \delta : L\mathcal{G} \to L\mathbf{G}_S \text{ an isomorphism of the associated } L\mathbf{G}\text{-torsors}\}/\cong . \end{aligned}$$

Here $(\mathcal{G}, \delta)$ and $(\mathcal{G}', \delta)$ are isomorphic if $\delta^{-1} \circ \delta'$ is an isomorphism $\mathcal{G}' \to \mathcal{G}$.

We fix a bound $[(R, \hat{Z}_R)]$ and a local $\mathbf{G}$-shtuka $\underline{\mathbb{G}} = (\mathbb{G}, \phi_{\mathbb{G}})$ over a field $k \in \mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}$ which is bounded by $\hat{Z}_R$. We consider the functor of bounded deformations of $\underline{\mathbb{G}}$ on the category of Artinian local $k[\![\zeta]\!]$-algebras with residue field k,

$$\begin{aligned} \mathrm{Def}_{\underline{\mathbb{G}}, \hat{Z}_R} : (\mathrm{Art}_k) \to & (\mathrm{Sets}) \\ A \mapsto & \{(\underline{\mathcal{G}}, \beta) \mid \underline{\mathcal{G}} \text{ a local } \mathbf{G}\text{-shtuka over Spec } A \text{ bounded by } \hat{Z}_R \\ & \quad \beta : \underline{\mathbb{G}} \to \underline{\mathcal{G}} \otimes_A k \text{ an isomorphism of local } \mathbf{G}\text{-shtukas}\}/\cong \end{aligned}$$

and require isomorphisms to be compatible with the maps β.

Remark 3-1. In Hartl and Viehmann [ibid.], 5 this functor and the deformation space representing it are explicitly described in case that $\mathbf{G}$ is split and $\hat{Z}_R$ associated with some $\omega \in \widetilde{W}$ as in Example 2-6, see below. This is generalized to reductive groups $\mathbf{G}$ in Viehmann and Wu [2016]. It would be interesting to have these results also in the more general context of parahoric group schemes.

Assume that $\mathbf{G}$ is a reductive group scheme over $\mathrm{Spec}\ \mathbb{F}_q[\![z]\!]$. Then the left hand side of Equation (2-7) is $W_0 \backslash \widetilde{W} / W_0$ where W_0 is the finite Weyl group of $\mathbf{G}_{\mathbb{F}_q}$. This double quotient is isomorphic to $X_*(T)_{\mathrm{dom}}$. Let $\mu \in X_*(T)_{\mathrm{dom}}$ and $(R, \hat{Z}_R)$ be the corresponding bound as in Example 2-6. Let $\underline{\mathbb{G}}$ be a local $\mathbf{G}$-shtuka bounded by μ over a field $k \in \mathrm{Nilp}_{\mathbb{F}_q[\![\zeta]\!]}$. Assume that there is a trivialization $\alpha : \underline{\mathbb{G}} \to (\mathbf{G}_k, b_0\sigma)$ for some $b_0 \in L\mathbf{G}(k)$. Using the boundedness, let $x \in Z_R(k)$ be the point defined by b_0^{-1}. By Hartl and Viehmann [2011], Thm. 5.6 and Viehmann and Wu [2016], Proposition 2.6 we have

Theorem 3-2. *Let* **G** *be reductive, and* $\hat{Z}_R$ *as in Example 2-6. Let* $\underline{\mathbb{G}}$ *and* α *be as above. Let* D *be the complete local ring of* $\hat{Z}_R$ *at the point* x. *It is a complete noetherian local ring over* $k[\![\zeta]\!]$. *Then* D *pro-represents the formal deformation functor* $\mathrm{Def}_{\underline{\mathbb{G}},\hat{Z}_R}$.

To study Newton strata in a family of local **G**-shtukas it is helpful to also have a corresponding stratification on deformations. However, for infinitesimal deformations such a question does not make sense. Instead we will replace the formal deformation space $\mathrm{Spf} D$ by $\mathrm{Spec}\, D$, using the following proposition. Here, a linearly topologized $\mathbb{F}_q[\![z]\!]$-algebra R is admissible if $R = \varprojlim R_\alpha$ for a projective system $(R_\alpha, u_{\alpha\beta})$ of discrete rings such that the filtered index-poset has a smallest element 0, all maps $R \to R_\alpha$ are surjective, and the kernels $I_\alpha := \ker u_{\alpha,0} \subset R_\alpha$ are nilpotent.

Proposition 3-3. *Let* R *be an admissible* $\mathbb{F}_q[\![z]\!]$*-algebra as above with filtered index-poset* $\mathbb{N}_0$. *Then the pullback under the natural morphism* $\mathrm{Spf} R \to \mathrm{Spec}\, R$ *defines a bijection between local* **G***-shtukas bounded by* μ *over* $\mathrm{Spec}\, R$ *and over* $\mathrm{Spf} R$.

Remark 3-4. Without a boundedness condition the pullback map is in general only injective. The corresponding result for p-divisible groups is shown by Messing and by de Jong. The above proposition is proved for split **G** in Hartl and Viehmann [2011], Proposition 3.16. However, the same proof also shows this assertion for all **G** that are reductive over $\mathrm{Spec}\, \mathbb{F}_q[\![z]\!]$.

4 Moduli spaces

In this section we fix **G**, an isogeny class of local **G**-shtukas and a bound $\hat{Z}$. We then define moduli spaces of local **G**-shtukas bounded by $\hat{Z}$ in the same way as Rapoport-Zink define their moduli spaces of p-divisible groups in Rapoport and Zink [1996].

Definition 4-1. Let $\underline{\mathbb{G}}_0$ be a local **G**-shtuka over $\mathbb{F}$. Let $\hat{Z} = [(R, \hat{Z}_R)]$ be a bound and denote its reflex ring by $R_{\hat{Z}} = \kappa[\![\xi]\!]$. It is a finite extension of $\mathbb{F}_q[\![\zeta]\!]$. Set $\breve{R}_{\hat{Z}} := \mathbb{F}[\![\xi]\!]$, and consider the functor

$$\begin{aligned}\breve{\mathfrak{M}} : (\mathrm{Nilp}_{\breve{R}_{\hat{Z}}})^\circ &\to (\mathrm{Sets})\\ S &\mapsto \{(\underline{\mathcal{G}}, \bar{\delta}) \mid \underline{\mathcal{G}} \text{ a local } \mathbf{G}\text{-shtuka over } S \text{ bounded by } \hat{Z}^{-1},\\ &\qquad \bar{\delta} : \underline{\mathcal{G}}_{\bar{S}} \to \underline{\mathbb{G}}_{0,\bar{S}} \text{ a quasi-isogeny}\}/\cong\end{aligned}$$

Here $\bar{S} := V_S(\zeta)$ is the zero locus of ζ in S, and two pairs $(\underline{\mathcal{G}}, \bar{\delta})$, $(\underline{\mathcal{G}}', \bar{\delta}')$ are isomorphic if $\bar{\delta}^{-1} \circ \bar{\delta}'$ lifts to an isomorphism $\underline{\mathcal{G}}' \to \underline{\mathcal{G}}$.

By Hartl and Viehmann [2017], 2.2 $\underline{\mathcal{G}}$ is bounded by $\hat{Z}^{-1}$ if and only if $\tau_{\mathcal{G}}^{-1}$ is bounded by $\hat{Z}_R$.

Let $\mathrm{QIsog}_{\mathbb{F}}(\underline{\mathbb{G}}_0)$ be the group of self-quasi-isogenies of $\underline{\mathbb{G}}_0$. It naturally acts on the functor $\breve{\mathfrak{M}}$. Since $\mathbb{F}$ has no non-trivial étale coverings, we may fix a trivialization $\underline{\mathbb{G}}_0 \cong ((L^+\mathbf{G})_{\mathbb{F}}, b\sigma^*)$ where $b \in L\mathbf{G}(\mathbb{F})$ represents the Frobenius morphism. Via such a trivialization, $\mathrm{QIsog}_{\mathbb{F}}(\underline{\mathbb{G}}_0)$ is identified with

$$J_b(\mathbb{F}_q((z))) := \{g \in G(\mathbb{F}((z))) \mid g^{-1}b\sigma(g) = b\}. \tag{4-2}$$

This is the set of $\mathbb{F}_q((z))$-valued points of an algebraic group J_b over $\mathbb{F}_q((z))$.

Theorem 4-3 (Arasteh Rad and Hartl [2014], Thm. 4.18 and Cor. 4.26, Hartl and Viehmann [2017], Rem. 3.5)**.** *The functor* $\breve{\mathfrak{M}}$ *is ind-representable by a formal scheme over* $\mathrm{Spf}\breve{R}_{\hat{Z}}$ *which is locally formally of finite type and separated. It is an ind-closed ind-subscheme of* $\mathrm{Flag}_{\mathbf{G}}\widehat{\times}_{\mathbb{F}_q}\mathrm{Spf}\breve{R}_{\hat{Z}}$.

By its analogy with moduli spaces of p-divisible groups, the formal scheme representing $\breve{\mathfrak{M}}$ is called a Rapoport-Zink space for bounded local $\mathbf{G}$-shtukas.

Let E be the quotient field of $R_{\hat{Z}}$, and $\breve{E} \cong \mathbb{F}((\xi))$ the completion of its maximal unramified extension. We write $\breve{\mathfrak{M}}^{\mathrm{an}}$ for the strictly $\breve{E}$-analytic space associated with $\breve{\mathfrak{M}}$.

Next we explain the construction of a tower of coverings of $\breve{\mathfrak{M}}^{\mathrm{an}}$. Roughly spoken, it is obtained by trivializing the (dual) Tate module of the universal local $\mathbf{G}$-shtuka over $\breve{\mathfrak{M}}^{\mathrm{an}}$.

Definition 4-4. Let S be an $\mathbb{F}((\zeta))$-scheme or a strictly $\mathbb{F}((\zeta))$-analytic space. Then an *étale local* $\mathbf{G}$*-shtuka* over S is a pair $\underline{\mathcal{G}} = (\mathcal{G}, \tau_{\mathcal{G}})$ consisting of an $L^+\mathbf{G}$-torsor $\mathcal{G}$ on S and an isomorphism $\tau_{\mathcal{G}} : \sigma^*\mathcal{G} \to \mathcal{G}$ of $L^+\mathbf{G}$-torsors.

Remark 4-5. Let $S = \mathrm{Spf}B$ be an affinoid admissible formal $\breve{R}_{\hat{Z}}$-scheme and S^{an} the associated strictly $\breve{E}$-analytic space. Consider a trivialized local $\mathbf{G}$-shtuka $((L^+\mathbf{G})_S, A\sigma^*)$ over S. Then $A \in \mathbf{G}(B[\![z]\!][\frac{1}{z-\zeta}])$. Note that

$$(z-\zeta)^{-1} = -\sum_{i=0}^{\infty} \zeta^{-i-1}z^i \in \mathcal{O}_{S^{\mathrm{an}}}(S^{\mathrm{an}})[\![z]\!]$$

implies $B[\![z]\!][\frac{1}{z-\zeta}] \subset \mathcal{O}_{S^{\mathrm{an}}}(S^{\mathrm{an}})[\![z]\!]$. Therefore $((L^+\mathbf{G})_S, A\sigma^*)$ induces an étale local $\mathbf{G}$-shtuka $((L^+\mathbf{G})_{S^{\mathrm{an}}}, A\sigma^*)$ over S^{an}.

To obtain similarly a universal family of étale local $\mathbf{G}$-shtukas over $\breve{\mathfrak{M}}^{\mathrm{an}}$, we cover $\breve{\mathfrak{M}}$ by affinoid admissible formal $\breve{R}_{\hat{Z}}$-schemes. For each of them one chooses a finite étale covering trivializing the local $\mathbf{G}$-shtuka, and applies the above construction. Descending the étale local $\mathbf{G}$-shtuka then yields the desired universal family, compare Hartl and Viehmann [ibid.], 6.

There are two approaches to construct the (dual) Tate module of an étale local $\mathbf{G}$-shtuka $\underline{\mathcal{G}}$ over a connected strictly $\breve{E}$-analytic space X. To define it directly, following Neupert

[2016], 2.6, consider for each $n \in \mathbb{N}$ the $L^+\mathbf{G}/\mathbf{G}_n$-torsor associated with $\mathcal{G}$ where $\mathbf{G}_n$ is the kernel of the projection $\mathbf{G}(\mathbb{F}_q[\![z]\!]) \to \mathbf{G}(\mathbb{F}_q[\![z]\!]/(z^n))$. The isomorphisms $\tau_{\mathcal{G}} : \sigma^*\mathcal{G} \to \mathcal{G}$ and $\sigma^* : \mathcal{G} \to \sigma^*\mathcal{G}$ then induce corresponding maps of $L^+\mathbf{G}/\mathbf{G}_n$-torsors. The invariants of $\tau \circ \sigma^*$ form a $\mathbf{G}(\mathbb{F}_q[\![z]\!]/(z^n))$-torsor which is trivialized by a finite étale covering of X. One can then define the Tate module of $\underline{\mathcal{G}}$ as the inverse limit over n of these torsors.

Alternatively, one can define it as a tensor functor, following Arasteh Rad and Hartl [2014], or Hartl and Viehmann [2017], 7: Fix a geometric base point $\bar{x}$ of X. We consider representations $\rho : \mathbf{G} \to \mathrm{GL}_r$ in $\mathrm{Rep}_{\mathbb{F}_q[\![z]\!]}(\mathbf{G})$. Let $\underline{M} = (M, \tau_M)$ be the étale local shtuka of rank r associated with $\rho_*\underline{\mathcal{G}}$ as in Example 2-2. Let $\underline{M}_{\bar{x}}$ denote its fiber over $\bar{x}$. The (dual) Tate module of $\underline{\mathcal{G}}$ with respect to ρ is the (dual) Tate module of $\underline{M}_{\bar{x}}$,

$$\check{T}\underline{M}_{\bar{x}} = \{m \in \underline{M}_{\bar{x}} \mid \tau_M(\sigma^* m) = m\},$$

a free $\mathbb{F}_q[\![z]\!]$-module of rank r. It carries a continuous monodromy action of $\pi_1^{\text{ét}}(X, \bar{x})$, which also factors through $\pi_1^{\text{alg}}(X, \bar{x})$. We obtain the *dual Tate module* of $\underline{\mathcal{G}}$ as a tensor functor

$$\check{T}_{\underline{\mathcal{G}},\bar{x}} : \mathrm{Rep}_{\mathbb{F}_q[\![z]\!]}\mathbf{G} \to \mathrm{Rep}^{\text{cont}}_{\mathbb{F}_q[\![z]\!]}(\pi_1^{\text{alg}}(X, \bar{x})).$$

Similarly, using rational representations in $\mathrm{Rep}_{\mathbb{F}_q(\!(z)\!)}\mathbf{G}$ one can define the *rational dual Tate module*

$$\check{V}_{\underline{\mathcal{G}},\bar{x}} : \mathrm{Rep}_{\mathbb{F}_q(\!(z)\!)}\mathbf{G} \to \mathrm{Rep}^{\text{cont}}_{\mathbb{F}_q(\!(z)\!)}(\pi_1^{\text{et}}(X, \bar{x})).$$

The two constructions of dual Tate modules are compatible in the sense that the tensor functor associated with the above torsor coincides with the tensor functor $\check{T}_{\mathcal{G},\bar{x}} : \mathrm{Rep}_{\mathbb{F}_q[\![z]\!]}\mathbf{G} \to (\mathbb{F}_q[\![z]\!]\text{-Loc})_X$ with values in the local systems of $\mathbb{F}_q[\![z]\!]$-lattices on X that can be associated with $\check{T}_{\underline{\mathcal{G}},\bar{x}}$ as in the second construction (compare Hartl and Viehmann [ibid.], Prop. 5.3).

We can now proceed to define level structures and the tower of coverings of $\breve{\mathfrak{M}}^{\text{an}}$. Let FMod_A denote the category of finite locally free A-modules. We consider the forgetful functors

$$\omega_A^\circ : \mathrm{Rep}_A\mathbf{G} \to \mathrm{FMod}_A \tag{4-6}$$

and

$$\mathit{forget} : \mathrm{Rep}_A^{\text{cont}}(\pi_1^{\text{et}}(X, \bar{x})) \to \mathrm{FMod}_A. \tag{4-7}$$

For an étale local **G**-shtuka $\underline{\mathcal{G}}$ over X the two sets

$$\mathrm{Triv}_{\underline{\mathcal{G}},\bar{x}}(\mathbb{F}_q[\![z]\!]) = \mathrm{Isom}^{\otimes}(\omega^\circ_{\mathbb{F}_q[\![z]\!]}, \mathit{forget} \circ \check{T}_{\underline{\mathcal{G}},\bar{x}})(\mathbb{F}_q[\![z]\!])$$
$$\mathrm{Triv}_{\underline{\mathcal{G}},\bar{x}}(\mathbb{F}_q(\!(z)\!)) = \mathrm{Isom}^{\otimes}(\omega^\circ_{\mathbb{F}_q(\!(z)\!)}, \mathit{forget} \circ \check{V}_{\underline{\mathcal{G}},\bar{x}})(\mathbb{F}_q(\!(z)\!))$$

are non-empty and carry natural actions of $\mathbf{G}(\mathbb{F}_q[\![z]\!]) \times \pi_1^{\mathrm{alg}}(X, \bar{x})$ and $G(\mathbb{F}_q(\!(z)\!)) \times \pi_1^{\mathrm{et}}(X, \bar{x})$, respectively. Here, the first factor acts via ω° and the second via the action on the Tate module. Note that non-emptiness needs our assumption that $\mathbf{G}$ has connected fibers, and thus slightly differs from the setting of Rapoport and Zink.

In the same way as in the arithmetic case we define coverings of $\breve{\mathfrak{M}}$.

Definition 4-8. Let $\underline{\mathcal{G}}$ be an étale local $\mathbf{G}$-shtuka over a connected $\breve{E}$-analytic space X, and let K be an open compact subgroup of $\mathbf{G}(\mathbb{F}_q[\![z]\!])$. An *integral K-level structure* on $\underline{\mathcal{G}}$ is a $\pi_1^{\mathrm{alg}}(X, \bar{x})$-invariant K-orbit in $\mathrm{Triv}_{\underline{\mathcal{G}},\bar{x}}(\mathbb{F}_q[\![z]\!])$.

For an open subgroup $K \subset \mathbf{G}(\mathbb{F}_q[\![z]\!])$ let X^K be the functor on the category of $\breve{E}$-analytic spaces over X parametrizing integral K-level structures on the local $\mathbf{G}$-shtuka $\underline{\mathcal{G}}$ over X.

Let $K \subset \mathbf{G}(\mathbb{F}_q(\!(z)\!))$ be compact open. Let $K' \subset K$ be a normal subgroup of finite index with $K' \subset \mathbf{G}(\mathbb{F}_q[\![z]\!])$. Then $gK' \in K/K'$ acts on $X^{K'}$ by sending the triple $(\underline{\mathcal{G}}, \delta, \eta K')$ over $X^{K'}$ to the triple $(\underline{\mathcal{G}}, \delta, \eta g K')$.

We define $\breve{\mathfrak{M}}^K$ as the $\breve{E}$-analytic space which is the quotient of $\breve{\mathfrak{M}}^{K'} := \breve{\mathfrak{M}}^{\mathrm{an},K'}$ by the finite group K/K'. It is independent of the choice of K'. In particular, $\breve{\mathfrak{M}}^{K_0} = (\breve{\mathfrak{M}})^{\mathrm{an}}$ for $K_0 = \mathbf{G}(\mathbb{F}_q[\![z]\!])$.

By Hartl and Viehmann [2017], Cor. 7.13 one can also describe $\breve{\mathfrak{M}}^K$ directly as a parameter space of local $\mathbf{G}$-shtukas with level structure.

Furthermore, if $K' \subseteq K$ is as above, the action of $gK' \in K/K'$ on $\breve{\mathfrak{M}}^{K'}$ can be described in this interpretation as a Hecke correspondence.

5 The geometry of the special fiber

We fix again a bound $(R, \hat{Z}_R)$ and let κ be the residue field of R. Let Z^{-1} be the special fiber of the inverse bound $\hat{Z}^{-1}$ over κ. The affine Deligne-Lusztig variety associated with an element $b \in L\mathbf{G}(\mathbb{F})$ and Z^{-1} is the reduced closed ind-subscheme $X_{Z^{-1}}(b) \subset \mathrm{Flag}_{\mathbf{G}}$ whose k-valued points (for any field extension k of $\mathbb{F}$) are given by

$$X_{Z^{-1}}(b)(k) = \{g \in \mathrm{Flag}_{\mathbf{G}}(k) \mid g^{-1}b\sigma(g) \in Z^{-1}(k)\}. \tag{5-1}$$

Our conditions on the bound imply that Z^{-1} is a left-$L^+\mathbf{G}$-invariant subscheme of $\mathrm{Flag}_{\mathbf{G}}$. The $L^+\mathbf{G}$-orbits in $\mathrm{Flag}_{\mathbf{G}}$ correspond bijectively to the elements of $\widetilde{W}^{\mathbf{G}} \backslash \widetilde{W} / \widetilde{W}^{\mathbf{G}}$, compare Example 2-6. For $x \in \widetilde{W}$ let

$$X_x(b)(k) = \{g \in \mathrm{Flag}_{\mathbf{G}}(k) \mid g^{-1}b\sigma(g) \in L^+\mathbf{G}xL^+\mathbf{G}\}. \tag{5-2}$$

Thus by left invariance every $X_{Z^{-1}}(b)(k)$ is a union of affine Deligne-Lusztig varieties of the form $X_x(b)$. By the boundedness condition on $\hat{Z}_R$, this union is finite. In the arithmetic context, such unions of affine Deligne-Lusztig varieties have been studied for example in He [2016].

Affine Deligne-Lusztig varieties are the underlying reduced subschemes of Rapoport-Zink spaces:

Theorem 5-3 (Hartl and Viehmann [2011], 6, Arasteh Rad and Hartl [2014], Thm. 4.18, Cor. 4.26)**.** *The underlying reduced subscheme of the moduli space* $\breve{\mathfrak{M}}$ *associated with* $\hat{Z}_R$ *and* b *as in Theorem 4-3 equals* $X_{Z^{-1}}(b)$. *It is a scheme locally of finite type and separated over* $\mathbb{F}$, *all of whose irreducible components are projective.*

Thus to describe the special fiber of the Rapoport-Zink spaces $\breve{\mathfrak{M}}$ it is enough to study the affine Deligne-Lusztig varieties $X_x(b)$. For an overview of the current state of the art of this field compare Viehmann [2015] and X. He's talk at this ICM He [2018]. We will here only discuss one class of results that were obtained as an application of the relation to local **G**-shtukas, and that were one of the initial goals of this theory. For the following result it is essential to assume that **G** is reductive over $\mathbb{F}_q[\![z]\!]$. For more general groups there are counterexamples to all assertions in the theorem. Recall that for reductive groups $\widetilde{W}^{\mathbf{G}}\backslash\widetilde{W}/\widetilde{W}^{\mathbf{G}} \cong X_*(T)_{\mathrm{dom}}$. Let $\mathrm{def}(b) = \mathrm{rk}\, G - \mathrm{rk}_{\mathbb{F}_q((z))} J_b$ where J_b is as in (4-2). Note that $X_*(T)_{\mathrm{dom}}$ is partially ordered by the Bruhat ordering. Via the induced ordering on the Newton points (requiring equality of Kottwitz points) also $B(G)$ inherits an ordering. Let l denote the length of a maximal chain between two elements in the partially ordered set $B(G)$, and let ρ be the half-sum of the positive roots of **G**.

Theorem 5-4 (Viehmann [2013], Hamacher and Viehmann [2017], 3.2)**.** *Let* $\mu_1 \preceq \mu_2 \in X_*(T)$ *be dominant coweights and* $[b] \in B(G)$. *Assume that* $\kappa_G(b) = \mu_2$ *in* $\pi_1(G)_\Gamma$ *and that* $\nu_b \preceq \mu_2$.

1. *Let*

$$S_{\mu_1,\mu_2} = \bigcup_{\mu_1 \preceq \mu' \preceq \mu_2} L^+\mathbf{G}\mu'(z)L^+\mathbf{G}.$$

Then the Newton stratum $\mathfrak{N}_{[b]} = [b] \cap S_{\mu_1,\mu_2}$ *is pure of codimension*

$$l([b],[\mu_2]) = \langle \rho, \mu_2 - \nu_b \rangle + \frac{1}{2}\mathrm{def}(b)$$

in S_{μ_1,μ_2}. *The closure of* $\mathfrak{N}_{[b]}$ *in* S_{μ_1,μ_2} *is the union of all* $\mathfrak{N}_{[b']}$ *for* $[b']$ *with* $\kappa_G(b') = \mu_2$ *in* $\pi_1(\mathbf{G})_\Gamma$ *and* $\nu_{b'} \preceq \nu_b$.

2. $X_{\mu_2}(b)$ *and* $X_{\preceq\mu_2}(b) = \bigcup_{\mu\preceq\mu_2} X_\mu(b)$ *are equidimensional of dimension*

$$\dim X_{\mu_2}(b) = \dim\ X_{\preceq\mu_2}(b) = \langle \rho, \mu_2 - \nu_b \rangle - \frac{1}{2}\mathrm{def}(b).$$

Here, the first assertion is most useful for $\mu_1 = \mu_2$, or for μ_1 the unique minuscule coweight with $\mu_1 \preceq \mu_2$.

Note that one needs to define the notions of codimension and of the closure of the infinite-dimensional schemes $\mathfrak{N}_{[b]}$. Both of these definitions use that there is an open subgroup H of $L\mathbf{G}^+$ such that the Newton point of an element of S_{μ_1,μ_2} only depends on its H-coset, an element of the finite-dimensional scheme $S_{\mu_1,\mu_2}/H$, compare Viehmann [2013], 4.3. These (co)dimensions are directly linked to the dimensions of Newton strata in the universal deformation of a local $\mathbf{G}$-shtuka as in Section 3. In fact, Theorem 5-4 is equivalent to a completely analogous result on dimensions and closure relations in that context, and this equivalence is also used in the proof of Theorem 5-4.

The proofs of the two parts of the theorem are closely linked: One first shows the upper bound on $\dim X_{\preceq\mu_2}(b)$ and how to compute $\dim \mathfrak{N}_{[b]}$ from $\dim X_{\preceq\mu_2}(b)$. Then a purity result for the Newton stratification proves the lower bound and equidimensionality. For the proof compare Hamacher and Viehmann [2017], 3.2 explaining how the proof for split $\mathbf{G}$ in Viehmann [2013] can be generalized. The theorem also inspired a corresponding theory for Newton strata in Shimura varieties of Hodge type by Hamacher.

6 Comparison to global $\mathfrak{G}$-shtukas

6.1 Global $\mathfrak{G}$-shtukas. Moduli spaces of bounded global $\mathfrak{G}$-shtukas are the function field analogue of Shimura varieties. Using these moduli spaces Lafforgue [2012] showed one direction of the global Langlands correspondence for all groups $\mathfrak{G}$ in the function field case. In this section we describe the relation between moduli spaces of bounded local $\mathbf{G}$-shtukas and global $\mathfrak{G}$-shtukas.

To define them, let C be a smooth projective geometrically irreducible curve over $\mathbb{F}_q$, and let $\mathfrak{G}$ be a parahoric group scheme over C, i.e. a smooth affine group scheme with connected fibers whose generic fiber is reductive over $\mathbb{F}_q(C)$ and such that for every point v of C for which the fiber above v is not reductive, the group scheme $\mathfrak{G}_v$ is a parahoric group scheme over A_v, as defined by Bruhat and Tits. Here A_v denotes the completion of $k[C]$ at v.

Definition 6-1. A *global shtuka* with n legs over a scheme S is a tuple $(\mathcal{G}, s_1, \dots, s_n, \varphi)$ where

1. $\mathcal{G} \in \mathcal{H}^1(C, \mathfrak{G})(S)$ is a $\mathfrak{G}$-torsor over $C \times_{\mathbb{F}_q} S$
2. the $s_i \in C(S)$ are pairwise disjoint S-valued points, called the legs
3. φ is an isomorphism

$$\varphi : \sigma^*\mathcal{G}|_{C\times_{\mathbb{F}_q} S\setminus\{\bigcup_i \Gamma_{s_i}\}} \to \mathcal{G}|_{C\times_{\mathbb{F}_q} S\setminus\{\bigcup_i \Gamma_{s_i}\}},$$

called the Frobenius isomorphism.

The stack $\nabla_n\mathcal{H}^1(C,\mathfrak{G})$ over $\operatorname{Spec}\mathbb{F}_q$ is the *moduli stack of global $\mathfrak{G}$-shtukas with n legs*.

There is a canonical morphism $\nabla_n\mathcal{H}^1(C,\mathfrak{G})\to C^n\setminus\Delta$ mapping a global $\mathfrak{G}$-shtuka to its legs.

We fix distinct places $c_i\in C(\mathbb{F})$ for $1\le i\le n$, and write $c=(c_1,\dots,c_n)$. Let A_c be the completion of the local ring $\mathcal{O}_{C^n,c}$, and let $\mathbb{F}_c$ be the residue field. Then $A_c\cong\mathbb{F}_c[\![\zeta_1,\dots,\zeta_n]\!]$.

Writing $A_{c_i}\cong\mathbb{F}_{c_i}[\![z]\!]$, let $\mathbf{G}_{c_i}=\mathfrak{G}\times_C\operatorname{Spec}A_{c_i}$, a parahoric group scheme over $\operatorname{Spec}\mathbb{F}_{c_i}[\![z]\!]$, and $\mathbf{G}_i:=\operatorname{Res}_{\mathbb{F}_{c_i}[\![z]\!]/\mathbb{F}_q[\![z]\!]}\mathbf{G}_{c_i}$.

Let

$$\nabla_n\mathcal{H}^1(C,\mathfrak{G})^c=\nabla_n\mathcal{H}^1(C,\mathfrak{G})\widehat{\times}_{C^n}\operatorname{Spf}A_c$$

be the formal completion of the stack $\nabla_n\mathcal{H}^1(C,\mathfrak{G})$ along c. It parametrizes global $\mathfrak{G}$-shtukas with n legs in the formal neighborhoods of the c_i. We want to define a global-local functor associating with such a global $\mathfrak{G}$-shtuka local shtukas at each of these places, compare Arasteh Rad and Hartl [2014], or Neupert [2016], 3.2. Let $(\mathcal{G},(s_i),\varphi)\in\nabla_n\mathcal{H}^1(C,\mathfrak{G})^c(S)$. In other words, $s_i:S\to C$ factors through $\operatorname{Spf}A_{c_i}$. We consider each place c_i separately. Let $S_{c_i}=\operatorname{Spec}\mathbb{F}_{c_i}\times_{\operatorname{Spec}\mathbb{F}_q}S$. The base change of $\mathcal{G}$ to $\operatorname{Spf}A_{c_i}\widehat{\times}_{\mathbb{F}_q}S=S_{c_i}[\![z]\!]$ defines an element of $\mathcal{H}^1(\mathbb{F}_q,\mathbf{G}_{c_i})(S_{c_i}[\![z]\!])$. We now use the following translation, cf. Hartl and Viehmann [2011], Prop. 2.2a, Neupert [2016], Prop. 3.2.4, or Arasteh Rad and Hartl [2014], 2.4.

Proposition 6-2. *Let S be a scheme over $\mathbb{F}_q$. There is a natural equivalence of categories*

$$\mathcal{H}^1(\mathbb{F}_q,\mathbf{G})(S[\![z]\!])\to\mathcal{H}^1(\mathbb{F}_q,L^+\mathbf{G})(S).$$

We thus obtain an element of

$$\mathcal{H}^1(\mathbb{F}_q,L^+\mathbf{G}_{c_i})(S_{c_i})=\mathcal{H}^1(\mathbb{F}_q,L^+\mathbf{G}_i)(S),$$

or an $L^+\mathbf{G}_i$-torsor $\mathcal{G}_i$ over S.

Also, the Frobenius morphism φ induces local Frobenius morphisms φ_i on the $\mathbf{G}_i$-torsors for each i. Altogether we obtain a functor

$$\mathfrak{L}=(\mathfrak{L}_1,\dots,\mathfrak{L}_n):\nabla_n\mathcal{H}^1(C,\mathfrak{G})^c\to\prod_i Sht_{\mathbf{G}_i},\qquad(\mathcal{G},(s_i),\varphi)\mapsto((\mathcal{G}_i,\varphi_i))$$

called the global-local functor.

In the same way, one can associate with every global $\mathfrak{G}$-shtuka $\underline{\mathcal{G}}\in\nabla_n\mathcal{H}^1(C,\mathfrak{G})^c(S)$ for some scheme S and every fixed place $s\in C\setminus\{c_i\}$ an étale local $\mathbf{G}_s$-shtuka $\mathfrak{L}_s(\underline{\mathcal{G}})=$

$\underline{\mathcal{G}}_s$. One can then define level structures on global $\mathfrak{G}$-shtukas away from the legs using the corresponding notion for local $\mathbf{G}$-shtukas. For a detailed discussion compare Neupert [2016], 3.4.

Definition 6-3. Let $\underline{\mathcal{G}} = (\mathcal{G}, (s_i), \varphi)$ be a global $\mathfrak{G}$-shtuka over S, let $D_0 \subset C$ be a finite reduced subscheme with $s_i \in (C \setminus D_0)(S)$ for every i. For every $v \in D_0$ fix an open subgroup $U_v \subseteq L^+G_v(\mathbb{F}_q)$ and let $U = \prod U_v$. Then an *integral U-level structure* on $\underline{\mathcal{G}}$ consists of an integral U_v-level structure of $\mathfrak{L}_v(\underline{\mathcal{G}})$ for every $v \in D_0$.

We denote by $\nabla \mathcal{H}^1_U(C, \mathfrak{G})^c$ the *stack of global $\mathfrak{G}$-shtukas* in $\nabla \mathcal{H}^1(C, \mathfrak{G})^c$ with U-level structure.

Rather ad hoc boundedness conditions are defined as follows (compare Arasteh Rad and Hartl [2013]).

Definition 6-4. Let $\hat{Z}_c = (\hat{Z}_i)_i$ be a tuple of bounds $\hat{Z}_i \subset \widehat{\mathrm{Flag}}_{\mathbf{G}_i}$. Then $(\mathcal{G}, (s_i), \varphi) \in \nabla_n \mathcal{H}^1(C, \mathfrak{G})^c(S)$ is *bounded by* $\hat{Z}_c$ if for every i the associated local $\mathbf{G}_i$-shtuka $(\mathcal{G}_i, \varphi_i)$ is bounded by $\hat{Z}_i$. We denote by $\nabla_n^{\hat{Z}_c} \mathcal{H}^1(C, \mathfrak{G})^c$ the substack of $\nabla_n \mathcal{H}^1(C, \mathfrak{G})^c$ of global $\mathfrak{G}$-shtukas bounded by $\hat{Z}_c$, and analogously for $\nabla_n^{\hat{Z}_c} \mathcal{H}^1_U(C, \mathfrak{G})^c$.

There are two other definitions of boundedness: A (seemingly more natural) one by Varshavsky [2004], Def. 2.4b for which, however, it is not clear how to establish a compatibility between bounded global and local shtukas. In Neupert [2016], 3.3 Neupert gives a global definition of boundedness, generalizing that of Varshavsky and he shows that it coincides with the one presented here.

6.2 The Serre-Tate theorem. Classically, the Serre-Tate theorem gives an equivalence of categories between deformations of an abelian variety over a scheme on which p is locally nilpotent, and its p-divisible group. By work of Arasteh Rad and Hartl, the analog in our situation also holds, taking into account that shtukas have several legs, and requiring them to be in the formal neighborhoods of fixed places of C.

Let S be in Nilp_{A_c} and let $j : \overline{S} \to S$ be a closed subscheme defined by a locally nilpotent sheaf of ideals. Let $c_1, \dots, c_n \in C$ and let $(\mathcal{G}, (s_i), \varphi) \in \nabla_n \mathcal{H}^1(C, \mathfrak{G})^c(\overline{S})$. Let $((\mathcal{G}_i, \varphi_i)) = \mathfrak{L}(\mathcal{G}, (s_i), \varphi)$.

Let $\mathrm{Def}_S((\mathcal{G}, (s_i), \varphi))$ be the category of lifts of $(\mathcal{G}, (s_i), \varphi)$ to S, i.e. of pairs $(\underline{\mathcal{H}}, \alpha)$ where $\underline{\mathcal{H}} \in \nabla_n \mathcal{H}^1(C, \mathfrak{G})^c(S)$ and where α is an isomorphism between $(\mathcal{G}, (s_i), \varphi)$ and the base change of $\underline{\mathcal{H}}$ to $\overline{S}$. Morphisms of lifts are isomorphisms compatible with the morphisms α. For the local $\mathbf{G}_i$-shtukas we define analogously a category of lifts $\mathrm{Def}_S((\mathcal{G}_i, \varphi_i))$.

Theorem 6-5 (Arasteh Rad and Hartl [2014], Th. 5.10)**.** *Let $(c_i)_i \in C^n$, let $(\mathcal{G}, (s_i), \varphi) \in \nabla_n \mathcal{H}^1(C, \mathfrak{G})^c(\overline{S})$ and let $((\mathcal{G}_i, \varphi_i)) = \mathfrak{L}(\mathcal{G}, (s_i), \varphi)$. Then the global-local functor induces an equivalence of categories* $\mathrm{Def}_S((\mathcal{G}, (s_i), \varphi)) \to \prod_i \mathrm{Def}_S((\mathcal{G}_i, \varphi_i))$.

6.3 Foliations. We begin by considering the uniformization morphism for Newton strata by Rapoport-Zink spaces, following Neupert [2016], 5 and Arasteh Rad and Hartl [2014], 5, and paralleling the uniformization result for p-divisible groups of Rapoport and Zink [1996]. In this subsection we assume that $\mathfrak{G}$ is the base change to C of a reductive group G over $\mathbb{F}_q$. It would be interesting to have a generalization of this result to the more general context considered before.

Let T denote a maximal torus of G, and fix a Borel subgroup B containing T. Let C be as above and fix characteristic places $c_1, \dots, c_n \in C$. Let $\mu_i \in X_*(T)$ be dominant cocharacters defined over a finite extension E of $\mathbb{F}_q$, and $\mu = (\mu_1, \dots, \mu_n) \in X_*(T)^n$. Fix decent local $\mathbf{G}_i$-shtukas $\underline{\mathbb{G}_i} = (L^+\mathbf{G}_{iE}, b_i\sigma^*)$ over Spec E, in particular $b_i \in L\mathbf{G}_i(E)$. Let $\mathfrak{M}_{b_i}^{\preceq\mu_i}$ be the Rapoport-Zink space associated with $[b_i]$ and the bound given by μ_i.

Theorem 6-6 (Neupert [2016], Theorem 5.1.18, Arasteh Rad and Hartl [2013], Theorem 7.4)**.** *Let S be a DM-stack over* $\operatorname{Spf} E[\![\zeta_1, \dots, \zeta_n]\!]$ *such that* $(\zeta_1, \dots, \zeta_n)$ *is locally nilpotent on S. Let $(\mathscr{G}_0, \phi_0, \psi_0) \in \nabla_n \mathscr{H}_U^1(C, \mathfrak{G})^c(S)$ for some congruence subgroup U. For each place c_i assume that there is, and fix, an isomorphism $\mathfrak{L}_{c_i}(\mathscr{G}_0, \phi_0) \cong \underline{\mathbb{G}_i}_S$. Then there is a morphism of formal DM-stacks over* $\operatorname{Spf} E[\![\zeta_1, \dots, \zeta_n]\!]$

$$S \times_{\operatorname{Spf} E[\![\zeta_1,\dots,\zeta_n]\!]} \prod_i \mathfrak{M}_{b_i}^{\preceq\mu_i} \to \nabla_n^{\mu} \mathscr{H}_U^1(C, \mathfrak{G})^c$$

which is ind-proper and formally étale.

This theorem has also generalizations to coverings of the moduli spaces associated with compatible level structures, and is then equivariant with respect to the action of $\mathfrak{G}(\mathbb{A}_{\mathbb{Q}}^{c_i})$ by Hecke correspondences.

We consider for the universal global $\mathfrak{G}$-shtuka $\underline{\mathscr{G}}$ over $\nabla_n^{\mu} \mathscr{H}_U^1(C, \mathfrak{G})^c$ the strata

$$\mathfrak{N}_{\underline{\mathscr{G}}}^{([b_i])_i} = \bigcap_i \mathfrak{N}_{[b_i], \mathfrak{L}_i(\underline{\mathscr{G}})} \subseteq \nabla_n^{\mu} \mathscr{H}_U^1(C, \mathfrak{G})^c.$$

Here, the $\mathfrak{N}_{[b_i], \mathfrak{L}_i(\underline{\mathscr{G}})}$ are the Newton strata in $\nabla_n^{\mu} \mathscr{H}_U^1(C, \mathfrak{G})^c$ associated with the local $\mathbf{G}_i$-shtuka $\mathfrak{L}_i(\underline{\mathscr{G}})$.

Then the uniformization morphism of Theorem 6-6 maps each geometric point of the left hand side to $\mathfrak{N}_{\underline{\mathscr{G}}}^{([b_i])_i}$ where the $[b_i]$ are the classes of the elements used to define the Rapoport-Zink space. Refining this description we now consider a foliation structure on these Newton strata, following Neupert [2016], 5.

In the arithmetic case, Oort and later Mantovan [2004] defined a foliation structure on Newton strata on Shimura varieties, by Rapoport-Zink spaces and so-called Igusa varieties. The latter are defined as covers of a central leaf, i.e. the locus in the Shimura variety where the p-divisible group is in a very particular isomorphism class. In our situation, the definition is slightly more involved. One needs to pass to the perfection of all involved moduli spaces to define the analog of Igusa varieties, and then to construct the foliation morphism.

Recall that a Newton stratum only depends on $L\mathbf{G}_i$-σ-conjugacy classes $[b_i]$, and not on individual representatives. A fundamental alcove in $[b_i]$ is an element x_{b_i} of $\widetilde{W}$ such that all (or one) representative $b_{i,0}$ in $N(L)$ is contained in $[b_i]$, and such that the length $\ell(x_{b_i})$ is minimal with that property, compare Viehmann [2014], Theorem 6.5, or Nie [2015]. One can then show that this length is equal to $\langle 2\rho_G, \nu_{b_i}\rangle$ where ρ_G is the half-sum of the positive roots of G and where ν_{b_i} is the dominant Newton point of $[b_i]$. Fundamental alcoves are a group-theoretic generalization of the minimal p-divisible groups studied by Oort. For each $[b_i]$ we fix such a representative $b_{i,0}$. The central leaf

$$\mathcal{C}_U^{(b_{i,0})} \subseteq \mathcal{N}_{\underline{\mathcal{G}}}^{([b_i])_i}$$

is then defined as the locus where the associated local $\mathbf{G}_i$-shtukas are isomorphic to $((L^+\mathbf{G}_i)_k, b_{i,0}$
It depends on the choice of the fundamental alcove, but not on that of the representative in $N(L)$.

We now define formal versions of these subschemes of $\nabla_n^{\mu}\mathcal{H}_U^1(C,G)$.

Definition 6-7. Recall that the formal completion of $C^n \setminus (c_1,\dots,c_n)$ is isomorphic to $\mathrm{Spf} A_c = \mathrm{Spf}\, \mathbb{F}_c[\![\zeta_1,\dots,\zeta_n]\!]$.

Let $\mathfrak{N}_U^{([b_i])_i}$ denote the formal completion of $\nabla_n^{\mu}\mathcal{H}_U^1(C,G)^c$ along the Newton stratum $\mathcal{N}_{\underline{\mathcal{G}}}^{([b_i])_i}$. This is not $(\zeta_1,\dots,\zeta_n)$-adic.

Let $\mathfrak{C}_U^{(b_{i,0})}$ be the locus in $\nabla_n^{\mu}\mathcal{H}_U^1(C,G)^c$ such that after an fpqc-covering each local $\mathbf{G}_i$-shtuka associated with the universal global $\mathfrak{G}$-shtuka is isomorphic to $(L^+\mathbf{G}_i, b_{i,0}\sigma^*)$. One can then show that this is represented by a $(\zeta_1,\dots,\zeta_n)$-adic formal scheme over $\mathrm{Spf} A_c$, called the *central leaf.* (Neupert [2016], Prop. 6.1.7)

By $X^\sharp$ we denote the perfection of a formal scheme X. It is again a formal scheme.

For $d \in \mathbb{N}$ we consider the subgroups $K_d = \{g \in L^+\mathbf{G}_i \mid g \equiv 1 \pmod{z^{d+1}}\}$ and

$$I_d(b_{i,0}) = \bigcap_{N \geq 0} \phi^N(K_d)$$

where $\phi(g) = b_{i,0}^{-1}\sigma^{-1}(g)b_{i,0}$.

The following theorem shows existence of the formal Igusa varieties.

Theorem 6-8 (Neupert [ibid.], Cor. 6.1.9)**.** *For all tuples $(d_i)_i$, define $\mathfrak{Ig}_U^{(d_i)\sharp}(T)$ as the set of pairs consisting of an element of $\mathfrak{C}_U^{(b_i)\sharp}(T)$ together with I_{d_i}-truncated isomorphisms between the associated local G_i-shtukas and $(L^+G_i, b_{i,0}\sigma^*)$ for all i. Then this* *defines a sheaf* $\mathfrak{Ig}_U^{(d_i)\sharp}$ *over* $\mathfrak{C}_U^{(b_i)\sharp}$ *representable by a* $(\zeta_1, \dots, \zeta_n)$*-adic formal scheme which is finite étale over* $\mathfrak{C}_U^{(b_i)\sharp}$*. It is called a* formal Igusa variety *of level* (d_i)*.*

Similarly, $\mathfrak{Ig}_U^{(\infty_i)\sharp}$ *parametrizing trivializations of the whole local* G_i*-shtukas is representable by a formal scheme, isomorphic to* $\varprojlim \mathfrak{Ig}_U^{(d_i)\sharp}$*.*

The foliation structure is then established by the following result.

Theorem 6-9 (Neupert [ibid.], Thm. 6.2.1)**.** *There is a natural formally étale morphism of formal schemes over* $\operatorname{Spf} A_c^\sharp$

$$\hat{\pi}_{(\infty_i)} : \prod_i \mathfrak{M}_{b_{i,0}}^{\preceq \mu_i \sharp} \times_{\operatorname{Spf} A_c^\sharp} \mathfrak{Ig}_U^{(\infty_i)\sharp} \to \mathfrak{N}_U^{[b_i]\sharp}.$$

There are also variants of this result for the special fiber of the spaces, using bounds on the quasi-isogenies, and for the associated adic spaces in the sense of Huber.

The very rough idea of the construction of this morphism is the following: A point in the Igusa variety gives us a global shtuka, such that at each place c_i the associated local $\mathbf{G}_i$-shtuka is the one defined by $b_{i,0}$. Then a point in the Rapoport-Zink space $\mathfrak{M}_{b_{i,0}}^{\preceq \mu_i}$ gives us a modification by a quasi-isogeny of this local $\mathbf{G}_i$-shtuka, and we define a new global $\mathfrak{G}$-shtuka by keeping the old one away from the c_i and replacing it in a formal neighborhood of c_i by this new local $\mathbf{G}_i$-shtuka.

Applying the theorem to the cohomology of the moduli space of global $\mathfrak{G}$-shtukas, and its decomposition into Newton strata, one can express the cohomology of Newton strata, and therefore of the whole moduli space of global $\mathfrak{G}$-shtukas in terms of the cohomology of Rapoport-Zink moduli spaces and of Igusa varieties. For technical reasons the statement is only known over a finite extension E' of $\mathbb{F}_q$ that can be described in terms of $\mathfrak{G}$ and μ. Denote by K the function field of the curve C, by $\overline{K}$ an algebraic closure, and by $\Gamma_{E'}$ the absolute Galois group of E'.

Theorem 6-10 (Neupert [ibid.], Main Theorem 2)**.** *Let* $\nabla_n^\mu \mathscr{H}^1(C, \mathfrak{G})$ *be a moduli space of global* $\mathfrak{G}$*-shtukas, such that all connected components of* $\nabla_n^\mu \mathscr{H}^1(C, \mathfrak{G}) \times_{C^n \setminus \Delta} \operatorname{Spec} E'[\![\zeta_1, \dots, \zeta_n]\!]$ *are proper over* $\operatorname{Spec} E'[\![\zeta_1, \dots, \zeta_n]\!]$.
Then there exists a canonical isomorphism between the virtual $\mathfrak{G}(\mathbb{A}_{\mathbb{Q}}^{c_i}) \times \Gamma_{E'}$*-representations*

$$\sum_j (-1)^j H_c^j \left(\nabla_n^\mu \mathscr{H}^1(C, \mathfrak{G}) \times \overline{K}, \mathbb{Q}_\ell\right)$$

and

$$\sum_{[b_i]} \sum_{d,e,f} (-1)^{d+e+f} \mathrm{Tor}_d^{\mathcal{H}(\prod J_{b_{i,0}})} \Big(H_c^e \big(\prod \mathfrak{M}_{b_{i,0}}^{\preceq \mu_i} \times \mathbb{F}, R\Psi_\eta^{\mathrm{an}} \mathbb{Q}_\ell\big), \varinjlim_U \varinjlim_{d_i} H_c^f \big(\mathrm{Ig}_U^{(d_i)} \times \mathbb{F}, R\Psi_\eta^{\mathrm{an}} \mathbb{Q}_\ell\big)\Big).$$

7 Period spaces and the period map

In our context, period spaces are constructed as strictly $\mathbb{F}_q((\zeta))$-analytic spaces in the sense of Berkovich. In the generality presented here, they are introduced and studied in Hartl and Viehmann [2017]. We allow more general bounds than those associated with minuscule coweights. Even if the bound is given by a coweight as in Example 2-6, the bound is in general a union of Schubert cells, and not a single one as in the minuscule case. For these two reasons the period spaces have to be defined as subspaces of an affine Grassmannian instead of a (classical) flag variety. To define them, we consider the affine Grassmannian $\mathrm{Gr}_{\mathbf{G}}^{B_{\mathrm{dR}}}$ of Remark 2-5. The space of Hodge-Pink $\mathbf{G}$-structures bounded by $\hat{Z}$ is defined as $\mathcal{H}_{\mathbf{G},\hat{Z}} := \hat{Z}_E$. It is a projective subscheme of $\mathrm{Gr}_G^{B_{\mathrm{dR}}} \widehat{\times}_{\mathbb{F}_q((\zeta))} E$. Fix a local $\mathbf{G}$-shtuka $\underline{\mathbb{G}}_0$ over $\mathbb{F}$ and a trivialization $\underline{\mathbb{G}}_0 \cong (L^+\mathbf{G}_{\mathbb{F}}, b\sigma^*)$.

A z-isocrystal over $\mathbb{F}$ is a pair (D, τ_D) consisting of a finite-dimensional $\mathbb{F}((z))$-vector space D and an isomorphism $\tau_D : \sigma^* D \to D$ of $\mathbb{F}((z))$-vector spaces. A Hodge-Pink structure on (D, τ_D) over a field extension L of $\mathbb{F}((\zeta))$ is a free $L[\![z-\zeta]\!]$-submodule $\mathfrak{q}_D \subset D \otimes_{\mathbb{F}((z))} L((z-\zeta))$ of full rank, compare Hartl and Kim [2016], 5.

Definition 7-1. Let L and b be as above and let $\gamma \in \mathrm{Gr}_G^{B_{\mathrm{dR}}}(L)$. Let $\rho : G_{\mathbb{F}_q((z))} \to \mathrm{GL}_{n,\mathbb{F}_q((z))}$ be a representation and let V be the representation space. Then we associate with b (and ρ) the z-isocrystal $(V \otimes_{\mathbb{F}_q((z))} \mathbb{F}((z)), \rho(\sigma^* b)\sigma^*)$ over $\mathbb{F}$. With γ we associate the Hodge-Pink structure over L defined by

$$\mathfrak{q}_D(V) = \rho(\gamma) \cdot V \otimes_{\mathbb{F}_q((z))} L[\![z-\zeta]\!] \subset D \otimes_{\mathbb{F}((z))} L((z-\zeta)).$$

Let

$$\underline{D}_{b,\gamma}(V) = (V \otimes_{\mathbb{F}_q((z))} \mathbb{F}((z)), \rho(\sigma^* b)\sigma^*, \mathfrak{q}_D(V)).$$

Consider the Newton point $\nu_b \in X_*(T)$ as a homomorphism $\mathbb{D}_{\mathbb{F}((z))} \to G_{\mathbb{F}((z))}$. Recall the Kottwitz map $\kappa_G : G(\mathbb{F}((z))) \to \pi_1(G)_\Gamma$. We then define the *Hodge degree* and the *Newton degree* as

$$t_H(\underline{D}_{b,\gamma}(V)) = \rho_*(\kappa_G(\gamma)) \qquad \text{and} \qquad t_N(\underline{D}_{b,\gamma}(V)) = \det{}_V \circ \rho \circ \nu_b.$$

Using $\pi_1(\mathrm{GL}(V))_\Gamma = \mathbb{Z}$ and $\mathrm{Hom}(\mathbb{D}_{\mathbb{F}((z))}, \mathbb{G}_m) = \mathbb{Q}$ we can view these as rational numbers. (Compare Hartl and Viehmann [2017], 4 for a definition in terms of $\underline{D}_{b,\gamma}(V)$ only.)

We call $\underline{D}_{b,\gamma}(V)$ *weakly admissible* if the images of $[b]$ and γ in $\pi_1(G)_{\Gamma,\mathbb{Q}}$ coincide and for all subobjects the Hodge degree is less or equal to the Newton degree. Finally, the pair (b,γ) is weakly admissible if $\underline{D}_{b,\gamma}(V)$ is weakly admissible for a faithful representation ρ of G.

Let $\breve{\mathcal{H}}_{\mathbf{G},\hat{Z}} := \mathcal{H}_{\mathbf{G},\hat{Z}} \widehat{\times}_E \breve{E}$ where $\mathcal{H}_{\mathbf{G},\hat{Z}} = \hat{Z}_E$. The *period space* $\breve{\mathcal{H}}^{wa}_{\mathbf{G},\hat{Z},b}$ is then defined as the set of all γ in the associated $\breve{E}$-analytic space $\breve{\mathcal{H}}^{\mathrm{an}}_{\mathbf{G},\hat{Z}}$ such that (b,γ) is weakly admissible.

To define the admissible locus $\breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b} \subset \breve{\mathcal{H}}^{\mathrm{an}}_{\mathbf{G},\hat{Z}}$, one associates with $\underline{D}$ a pair of σ-bundles $\underline{\mathcal{E}}(\underline{D}) = \underline{\mathcal{E}}_{b,\gamma}(V)$ and $\underline{\mathcal{F}}(\underline{D}) = \underline{\mathcal{F}}_{b,\gamma}(V)$. The *admissible locus* is then defined as the subset over which $\underline{\mathcal{F}}_{b,\gamma}(V)$ has slope zero. Then $\breve{\mathcal{H}}^{wa}_{\mathbf{G},\hat{Z},b}$ and $\breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b}$ are open paracompact strictly $\breve{E}$-analytic subspaces of $\breve{\mathcal{H}}^{\mathrm{an}}_{\mathbf{G},\hat{Z}}$ (Hartl and Viehmann [ibid.], Theorem 4.20).

The neutral admissible locus $\breve{\mathcal{H}}^{na}_{\mathbf{G},\hat{Z},b} \subset \breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b}$ is defined by the condition that the images of $[b]$ and γ in $\pi_1(G)_\Gamma$ coincide.

We have the analog of the theorem of Colmez and Fontaine that "admissible implies weakly admissible", see Hartl [2011],Theorem 2.5.3. In other words, we have an inclusion

$$\breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b} \subseteq \breve{\mathcal{H}}^{wa}_{\mathbf{G},\hat{Z},b}.$$

Furthermore, we have $\breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b}(L) = \breve{\mathcal{H}}^{wa}_{\mathbf{G},\hat{Z},b}(L)$ for all finite field extensions $L/\breve{E}$. If L is algebraically closed, weakly admissible does *not* imply admissible in general.

For an $\mathbb{F}[\![\zeta]\!]$-algebra B, complete and separated with respect to a bounded norm $|\cdot| : B \to [0,1] \subset \mathbb{R}$ with $0 < |\zeta| < 1$, we consider the $\mathbb{F}((z))$-algebra

$$B[\![z,z^{-1}\} = \{\sum_{i\in\mathbb{Z}} b_i z^i \mid b_i \in B, |b_i|\,|\zeta|^{ri} \to 0\ (i \to -\infty) \text{ for all } r > 0\}$$

and the element $t_- = \prod_{i\in\mathbb{N}_0}(1 - \frac{\zeta^{q^i}}{z}) \in B[\![z,z^{-1}\}$.

We want to define the *period morphism*

$$\breve{\pi} : \breve{\mathfrak{M}}^{\mathrm{an}} \to \breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b}$$

as in Hartl and Viehmann [2017], 6 where $\mathfrak{M}$ is the Rapoport-Zink space associated with $\hat{Z}$ and b. Let S be an affinoid, strictly $\breve{E}$-analytic space and let $S \to \breve{\mathfrak{M}}^{\mathrm{an}}$ be a morphism of

$\breve{E}$-analytic spaces. With it we have to associate a morphism $S \to \breve{\mathcal{H}}^{\mathrm{an}}_{\mathbf{G},\hat{Z}}$. By construction of $\breve{\mathfrak{M}}^{\mathrm{an}}$ a morphism $S \to \breve{\mathfrak{M}}^{\mathrm{an}}$ is induced by a morphism from a quasi-compact admissible formal $\breve{R}_{\hat{Z}}$-scheme $\mathcal{S}$ with $\mathcal{S}^{\mathrm{an}} = S$ to $\breve{\mathfrak{M}}$. The latter corresponds to some $(\underline{\mathcal{G}}, \bar{\delta}) \in \breve{\mathfrak{M}}(\mathcal{S})$. After an étale covering $\mathcal{S}' = \mathrm{Spf}B' \to \mathcal{S}$ of admissible formal $\breve{R}_{\hat{Z}}$-schemes there is a trivialization $\alpha : \underline{\mathcal{G}}_{\mathcal{S}'} \cong ((L^+\mathbf{G})_{\mathcal{S}'}, A\sigma^*)$ for some $A \in \mathbf{G}(B'[\![z]\!][\frac{1}{z-\zeta}])$.

Denote by $\bar{\cdot}$ the reduction to $V(\zeta)$. Then the quasi-isogeny $\bar{\delta}$ induces via $\bar{\alpha}$ an element $\overline{\Delta} \in L\mathbf{G}(\overline{B}')$. One then shows that there is a uniquely determined element $\Delta \in \mathbf{G}(B'[\![z, z^{-1}\}[\frac{1}{t_-}])$ lifting $\overline{\Delta}$ with $\Delta A = b\sigma^*(\Delta)$ for A as above and b defining the Rapoport-Zink space. Consider the element

$$\gamma = \sigma^*(\Delta)A^{-1} \cdot \mathbf{G}(B'[\tfrac{1}{\zeta}][\![z-\zeta]\!]) \in \mathbf{G}(B'[\tfrac{1}{\zeta}](\!(z-\zeta)\!))/\mathbf{G}(B'[\tfrac{1}{\zeta}][\![z-\zeta]\!]).$$

The boundedness of $\underline{\mathcal{G}}$ then implies that γ factors through $\breve{\mathcal{H}}^{\mathrm{an}}_{\mathbf{G},\hat{Z}}$. It descends to a well-defined element of $\breve{\mathcal{H}}^{\mathrm{an}}_{\mathbf{G},\hat{Z}}(S)$ which is the image under the period morphism.

Proposition 7-2 (Hartl and Viehmann [2017], Prop. 6.9, 6.10)**.** *The period morphism factors through the open $\breve{E}$-analytic subspace $\breve{\mathcal{H}}^{a}_{G,\hat{Z},b}$ and induces an étale morphism*

$$\breve{\pi} : \breve{\mathfrak{M}}^{\mathrm{an}} \to \breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b}.$$

To describe the image of $\breve{\pi}$ recall the tensor functors of (4-6) and (4-7). The construction of the σ-bundle $\underline{\mathcal{F}}_{b,\gamma}(V)$ of Definition 7-1 has a generalization to a σ-bundle $\underline{\mathcal{F}}_b(V)$ over $\breve{\mathcal{H}}_{G,\hat{Z}}$ with fibers $\underline{\mathcal{F}}_{b,\gamma}(V)$. It further induces a canonical local system $\underline{\mathcal{V}}_b(V)$ of $\mathbb{F}_q(\!(z)\!)$-vector spaces on $\breve{\mathcal{H}}^{a}_{G,\hat{Z},b}$ with $\underline{\mathcal{V}}_b(V)_{\bar{\gamma}} = \bar{\gamma}^*\underline{\mathcal{F}}_b(V)^{\tau}$ where we take invariants under the isomorphism τ defining the σ-bundle. Let $\omega_{b,\bar{\gamma}} : \mathrm{Rep}_{\mathbb{F}_q(\!(z)\!)}G \to \mathrm{FMod}_{\mathbb{F}_q(\!(z)\!)}$ be the fiber functor with $\omega_{b,\bar{\gamma}}(V) := \underline{\mathcal{V}}_b(V)_{\bar{\gamma}} = \bar{\gamma}^*\underline{\mathcal{F}}_b(V)^{\tau}$. Then, one of the main results of Hartl and Viehmann [ibid.] is the following description of the image of the period map. The condition that G be unramified can also be replaced by another, more technical, but possibly more general condition.

Theorem 7-3. *Assume that G is unramified.*

1. *The image $\breve{\pi}(\breve{\mathfrak{M}}^{\mathrm{an}})$ of the period morphism is equal to the union of those connected components of $\breve{\mathcal{H}}^{a}_{\mathbf{G},\hat{Z},b}$ on which there is an $\mathbb{F}_q(\!(z)\!)$-rational isomorphism $\beta : \omega^{\circ} \to \omega_{b,\bar{\gamma}}$.*

2. *The rational dual Tate module $\check{V}_{\underline{\mathcal{G}}}$ of the universal local $\mathbf{G}$-shtuka $\underline{\mathcal{G}}$ over $\breve{\mathfrak{M}}^{\mathrm{an}}$ descends to a tensor functor $\check{V}_{\underline{\mathcal{G}}}$ from $Rep_{\mathbb{F}_q(\!(z)\!)}G$ to the category of local systems of $\mathbb{F}_q(\!(z)\!)$-vector spaces on $\breve{\pi}(\breve{\mathfrak{M}}^{\mathrm{an}})$. It carries a canonical $J_b(\mathbb{F}_q(\!(z)\!))$-linearization and is canonically $J_b(\mathbb{F}_q(\!(z)\!))$-equivariantly isomorphic to the tensor functor $\underline{\mathcal{V}}_b$.*

3. *The tower of strictly $\breve{E}$-analytic spaces $(\breve{\mathfrak{M}}^K)_{K\subset \mathbf{G}(\mathbb{F}_q((z)))}$ is canonically isomorphic over $\breve{\pi}(\breve{\mathfrak{M}}^{\mathrm{an}})$ in a Hecke and $J_b(\mathbb{F}_q((z)))$-equivariant way to the tower of étale covering spaces of $\breve{\pi}(\breve{\mathfrak{M}}^{\mathrm{an}})$ that is naturally associated with the tensor functor $\underline{\mathcal{V}}_b$.*

References

Greg W. Anderson (1993). "On Tate modules of formal t-modules". *IMRN* 2, pp. 41–52. MR: 1203253 (cit. on p. 1445).

Esmail Arasteh Rad and Urs Hartl (2013). "Uniformizing the moduli stacks of global G-Shtukas". To appear in IMRN. arXiv: 1302.6351 (cit. on pp. 1456, 1457).

– (2014). "Local $\mathbb{P}$-shtukas and their relation to global $\mathfrak{G}$-shtukas". *Münster J. Math.* 7.2, pp. 623–670. MR: 3426233 (cit. on pp. 1445, 1447, 1450, 1451, 1453, 1455–1457).

François Bruhat and Jacques Tits (1972). "Groupes réductifs sur un corps local: I. Données radicielles valuées". *Inst. Hautes Études Sci. Publ. Math.* 41, pp. 5–251. MR: 0327923 (cit. on p. 1444).

Alain Genestier (1996). "Espaces symétriques de Drinfeld". *Astérisque* 234, pp. ii+124. MR: 1393015 (cit. on p. 1445).

Thomas Haines and Michael Rapoport (2008). "On parahoric subgroups". appendix to Pappas and Rapoport [2008] (cit. on p. 1444).

Paul Hamacher and Eva Viehmann (Feb. 2017). "Irreducible components of minuscule affine Deligne-Lusztig varieties". arXiv: 1702.08287 (cit. on pp. 1453, 1454).

Urs Hartl (2011). "Period spaces for Hodge structures in equal characteristic". *Ann. of Math. (2)* 173.3, pp. 1241–1358. MR: 2800715 (cit. on p. 1461).

Urs Hartl and Wansu Kim (2016). "Local Shtukas, Hodge-Pink Structures and Galois Representations". *to appear in Proceedings of the conference on "t-motives: Hodge structures, transcendence and other motivic aspects", BIRS, Banff, Canada 2009, eds. G. Böckle, D. Goss, U. Hartl, M. Papanikolas, EMS* (cit. on p. 1460).

Urs Hartl and Eva Viehmann (2011). "The Newton stratification on deformations of local **G**-shtukas". *J. Reine Angew. Math.* 656, pp. 87–129. MR: 2818857 (cit. on pp. 1445, 1448, 1449, 1453, 1455).

– (2012). "Foliations in deformation spaces of local **G**-shtukas". *Adv. Math.* 229.1, pp. 54–78. MR: 2854170 (cit. on p. 1445).

– (Dec. 2017). "The generic fiber of moduli spaces of bounded local G-shtukas". arXiv: 1712.07936 (cit. on pp. 1447, 1449–1452, 1460–1462).

Xuhua He (2016). "Kottwitz-Rapoport conjecture on unions of affine Deligne-Lusztig varieties". *Ann. Sci. Éc. Norm. Supér. (4)* 49.5, pp. 1125–1141. arXiv: 1408.5838. MR: 3581812 (cit. on p. 1453).

Xuhua He (2018). "Some results on affine Deligne–Lusztig varieties". In: *Proc. of the International Congress of Mathematicians, Rio de Janeiro 2018*. Ed. by Boyan Sirakov, Paulo Ney de Souza, and Marcelo Viana. In this volume. World Scientific, pp. 1345–1366 (cit. on p. 1453).

Robert E. Kottwitz (1985). "Isocrystals with additional structure". *Compositio Math.* 56.2, pp. 201–220. MR: 809866 (cit. on p. 1445).

– (1997). "Isocrystals with additional structure. II". *Compositio Math.* 109.3, pp. 255–339. MR: 1485921 (cit. on p. 1445).

Vincent Lafforgue (Sept. 2012). "Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale". arXiv: 1209.5352 (cit. on p. 1454).

Elena Mantovan (2004). "On certain unitary group Shimura varieties". *Astérisque* 291. Variétés de Shimura, espaces de Rapoport-Zink et correspondances de Langlands locales, pp. 201–331. MR: 2074715 (cit. on p. 1457).

Stephan Neupert (2016). "Foliations and the cohomology of moduli spaces of bounded global G-shtukas". PhD thesis. TU Munich. arXiv: 1610.05935 (cit. on pp. 1445, 1450, 1455–1459).

Sian Nie (2015). "Fundamental elements of an affine Weyl group". *Math. Ann.* 362.1-2, pp. 485–499. MR: 3343886 (cit. on p. 1458).

George Pappas and Michael Rapoport (2008). "Twisted loop groups and their affine flag varieties". *Adv. Math.* 219.1. With an appendix by T. Haines and Rapoport, pp. 118–198. MR: 2435422 (cit. on pp. 1446, 1463).

Michael Rapoport and Melanie Richartz (1996). "On the classification and specialization of F-isocrystals with additional structure". *Compositio Math.* 103.2, pp. 153–181. MR: 1411570 (cit. on p. 1446).

Michael Rapoport and Thomas Zink (1996). *Period spaces for p-divisible groups*. Vol. 141. Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, pp. xxii+324. MR: 1393439 (cit. on pp. 1449, 1457).

Timo Richarz (2016). "Affine Grassmannians and geometric Satake equivalences". *IMRN* 2016, pp. 3717–3767. MR: 3544618 (cit. on p. 1446).

Yakov Varshavsky (2004). "Moduli spaces of principal F-bundles". *Selecta Math. (N.S.)* 10.1, pp. 131–166. MR: 2061225 (cit. on p. 1456).

Eva Viehmann (2013). "Newton strata in the loop group of a reductive group". *Amer. J. Math.* 135.2, pp. 499–518. MR: 3038719 (cit. on pp. 1453, 1454).

– (2014). "Truncations of level 1 of elements in the loop group of a reductive group". *Ann. of Math. (2)* 179.3, pp. 1009–1040. MR: 3171757 (cit. on p. 1458).

– (Nov. 2015). "On the geometry of the Newton stratification". arXiv: 1511.03156 (cit. on p. 1453).

Eva Viehmann and Han Wu (June 2016). "Central leaves in loop groups". To appear in Math. Res. Letters. arXiv: 1606.09444 (cit. on p. 1448).

Jean-Pierre Wintenberger (1997). "Propriétés du groupe tannakien des structures de Hodge p-adiques et torseur entre cohomologies cristalline et étale". *Ann. Inst. Fourier (Grenoble)* 47.5, pp. 1289–1334. MR: 1600395.

Received 2017-11-30.

Eva Viehmann
Fakultät für Mathematik
Technische Universität München – M11
Boltzmannstr. 3
85748 Garching
Germany
viehmann@ma.tum.de

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 2 (1465–1494)

HITCHIN TYPE MODULI STACKS IN AUTOMORPHIC REPRESENTATION THEORY

Zhiwei Yun (恽之玮)

Abstract

In the study of automorphic representations over a function field, Hitchin moduli stack and its variants naturally appear and their geometry helps the comparison of trace formulae. We give a survey on applications of this observation to a relative fundamental lemma, the arithmetic fundamental lemma and to the higher Gross-Zagier formula.

1 Introduction

1.1 Hitchin's original construction. In an influential paper by Hitchin [1987], he introduced the famous integrable system, the moduli space of Higgs bundles. Let X be a smooth proper and geometrically connected curve over a field k. Let G be a connected reductive group over k. Let $\mathcal{L}$ be a line bundle over X. An *$\mathcal{L}$-twisted G-Higgs bundle* over X is a pair $(\mathcal{E}, \varphi)$ where $\mathcal{E}$ is a principal G-bundle over X and φ is a global section of the vector bundle $\mathrm{Ad}(\mathcal{E}) \otimes \mathcal{L}$ over X. Here, $\mathrm{Ad}(\mathcal{E})$ is the vector bundle associated to $\mathcal{E}$ and the adjoint representation of G. The moduli stack $\mathcal{M}_{G,\mathcal{L}}$ of $\mathcal{L}$-twisted Higgs G-bundles over X is called the Hitchin moduli stack. Hitchin defined a map

$$f : \mathcal{M}_{G,\mathcal{L}} \to \mathcal{A}_{G,\mathcal{L}} \tag{1-1}$$

to some affine space $\mathcal{A}_{G,\mathcal{L}}$ by collecting invariants of φ such as its trace and determinant in the case $G = \mathrm{GL}_n$. The map f is called the *Hitchin fibration*. When $\mathcal{L} = \omega_X$ is the line bundle of 1-forms on X, Hitchin showed that f exhibited the stable part of $\mathcal{M}_{G,\omega_X}$ as a completely integrable system. He also gave concrete descriptions of the fibers of f in terms of spectral curves.

The works presented here have been supported by the NSF grants DMS-0969470/1302071/1736600 and the Packard Foundation.
MSC2010: primary 11F70; secondary 14D24, 22E57, 11F67.

1.2 Applications in geometric representation theory. Although discovered in the context of Yang-Mills theory, Hitchin moduli stacks have subsequently played important roles in the development of geometric representation theory.

When $\mathcal{L} = \omega_X$, $\mathfrak{M}_{G,\omega_X}$ is essentially the total space of the cotangent bundle of the moduli stack Bun_G of G-bundles over X. Therefore the categories of twisted D-modules on Bun_G give quantizations of $\mathfrak{M}_{G,\omega_X}$. Beilinson and Drinfeld studied such quantizations and used them to realize part of the geometric Langlands correspondence (namely when the $\widehat{G}$-connection comes from an oper). This can be viewed as a global analogue of the Beilinson-Bernstein localization theorem. A related construction in positive characteristic was initiated by Braverman and Bezrukavnikov [2007] for GL_n and extended to any G by Chen and Zhu [2017]. Hitchin's integrable system also plays a key role in the work of Kapustin and Witten [2007] which ties geometric Langlands correspondence to quantum field theory.

Hitchin moduli stacks have also been used to construct representations of the double affine Hecke algebra, giving global analogues of Springer representations. See Yun [2011a], Yun [2012] and Oblomkov and Yun [2016].

1.3 Applications in automorphic representation theory. Ngô [2006] made the fundamental observation that point-counting on Hitchin fibers is closely related to orbital integrals that appear in the study of automorphic representations. This observation, along with ingenious technical work, allowed Ngô to prove the Lie algebra version of the Fundamental Lemma conjectured by Langlands and Shelstad in the function field case, see Ngô [2010].

Group versions of the Hitchin moduli stack were introduced by Ngô using Vinberg semigroups. They are directly related to the Arthur-Selberg trace formula, as we will briefly review in Section 2.1. See recent works by Ngô [2014], Bouthier [2017] and Bouthier, Ngô, and Sakellaridis [2016] for applications of group versions of the Hitchin moduli stack.

1.4 Contents of this report. This report will focus on further applications of variants of Hitchin moduli stacks to automorphic representation theory.

In Section 2, we explain, in heuristic terms, why Hitchin-type moduli stacks naturally show up in the study of Arthur-Selberg trace formula and more generally, relative trace formulae. A relative trace formula calculates the L^2-pairing of two distributions on the space of automorphic forms of G given by two subgroups. Such pairings, when restricted to cuspidal automorphic representations, are often related to *special values* of automorphic L-functions. In Section 2.3, we will elaborate on the relative trace formulae introduced by Jacquet and Rallis, for which the fundamental lemma was proved in Yun [2011b].

In Section 3, we point out a new direction initiated in the works Yun [n.d.], Yun and Zhang [2017] and Yun and Zhang [n.d.]. Drinfeld introduced the moduli stack of Shtukas as an analogue of Shimura varieties for function fields, which turns out to allow richer variants than Shimura varieties. Cohomology classes of these moduli of Shtukas generalize the notion of automorphic forms. In Section 3.1 we review the basic definitions of Shtukas, and discuss the spectral decomposition for the cohomology of moduli of Shtukas. In Section 3.2, we introduce Heegner-Drinfeld cycles on the moduli of G-Shtukas coming from subgroups H of G. The relative trace in the context of Shtukas is then defined in Section 3.3 as the intersection pairing of two Heegner-Drinfeld cycles. We believe that such pairings, when restricted to the isotypical component of a cuspidal automorphic representation, are often related to *higher derivatives* of automorphic L-functions. We then explain why Hitchin-type moduli stacks continue to play a key role in the Shtuka context, and what new geometric ingredients are needed to study relative trace formulae in this setting.

In Section 3.4-Section 3.5 we survey what has been proven in this new direction. In Section 3.4, we review Yun and Zhang [2017] and Yun and Zhang [n.d.], in which we obtain formulae relating higher derivatives of automorphic L-functions for PGL_2 and the intersection numbers of Heegner-Drinfeld cycles. Our results generalize the Gross-Zagier formula in the function field case. In Section 3.5 we discuss the analogue of the fundamental lemma in the Shtuka setting. This was originally conjectured by W.Zhang under the name *arithmetic fundamental lemma*. We state an extension of his conjecture for function fields involving higher derivatives of orbital integrals, and sketch our strategy to prove it.

Acknowledgments. I would like to thank my mentors for bringing me to the exciting crossroads of algebraic geometry, representation theory and number theory. The influences of R. MacPherson, P. Deligne, G. Lusztig, B. Gross, B.C. Ngô and R. Bezrukavnikov on my career are especially important. I also learned a lot from my collaborators, colleagues and fellow students, to whom I would like to express my deep gratitude.

2 Hitchin moduli stack and trace formulae

Throughout this article we fix a finite field $k = \mathbb{F}_q$. Let X be a smooth, projective and geometrically connected curve over k of genus g. Let $F = k(X)$ be the function field of X. The places of F can be identified with the set $|X|$ of closed points of X. Let $\mathbb{A}$ denote the ring of adèles of F. For $x \in |X|$, let $\mathcal{O}_x$ be the completed local ring of X at x, and F_x (resp. k_x) be its fraction field (resp. residue field). We also use ϖ_x to denote a uniformizer of $\mathcal{O}_x$.

In this section we work with the classical notion of automorphic forms for groups over function fields. We shall briefly review the Arthur-Selberg trace formula and the relative trace formulae, and explain why Hitchin type moduli stacks naturally show up in the study of these trace formulae.

2.1 Arthur-Selberg trace formula. The Arthur-Selberg trace formula is an important tool in the theory of automorphic representations. For a detailed introduction to the theory over a number field we recommend Arthur's article Arthur [2005]. Here we focus on the function field case. The idea that Hitchin moduli stacks give a geometric interpretation of Arthur-Selberg trace formula is due to B.C. Ngô. For more details, we refer to Ngô [2006] for the Lie algebra version, and Frenkel and Ngô [2011] for the group version.

We ignore the issue of convergence in the discussion (i.e., we pretend to be working with an anisotropic group G), but we remark that the convergence issue lies at the heart of the theory of Arthur and we are just interpreting the easy part of his theory from a geometric perspective.

2.1.1 The classical setup. Let G be a split connected reductive group over k and we view it as a group scheme over X (hence over F) by base change. Automorphic forms for G are $\mathbb{C}$-valued smooth functions on the coset space $G(F)\backslash G(\mathbb{A})$. Fix a Haar measure μ_G on $G(\mathbb{A})$. Let $\mathfrak{A}$ be the space of automorphic forms for G. For any smooth compactly supported function f on $G(\mathbb{A})$, it acts on $\mathfrak{A}$ by right convolution $R(f)$.

The Arthur-Selberg trace formula aims to express the trace of $R(f)$ on $\mathfrak{A}$ in two different ways: one as a sum over conjugacy classes of $G(F)$ (the geometric expansion) and the other as a sum over automorphic representations (the spectral expansion). The primitive form of the geometric expansion reads

$$\mathrm{Tr}(R(f),\mathfrak{A})'' ='' \sum_{\gamma\in G(F)/\sim} J_\gamma(f) \tag{2-1}$$

where γ runs over $G(F)$-conjugacy classes in $G(F)$, and $J_\gamma(f)$ is the orbital integral

$$J_\gamma(f)=\mathrm{vol}(G_\gamma(F)\backslash G_\gamma(\mathbb{A}),\mu_{G_\gamma})\int_{G_\gamma(\mathbb{A})\backslash G(\mathbb{A})} f(g^{-1}\gamma g)\frac{\mu_G}{\mu_{G_\gamma}}(g)$$

where μ_{G_γ} is any Haar measure on the centralizer $G_\gamma(\mathbb{A})$ of γ. We write the equality sign in quotation marks [1] to indicate that the convergence issue has been ignored in (2-1). We will give a geometric interpretation of the geometric expansion.

[1] Later we will use the same notation (equal signs in quotation marks) to indicate heuristic identities.

Fix a compact open subgroup $K = \prod_{x\in|X|} K_x \subset G(\mathbb{A})$ and assume $\mathrm{vol}(K, \mu_G) = 1$. Let $\mathcal{Q}_K = C_c(G(F)\backslash G(\mathbb{A})/K)$ on which the Hecke algebra $C_c(K\backslash G(\mathbb{A})/K)$ acts. For $g \in G(\mathbb{A})$, there is a Hecke correspondence attached to the double coset $KgK \subset G(\mathbb{A})$

$$G(F)\backslash G(\mathbb{A})/K \xleftarrow{p_0} G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1}) \xrightarrow{q_0} G(F)\backslash G(\mathbb{A})/K \tag{2-2}$$

where p_0 is the natural projection and q_0 is induced by right multiplication by g. The action of $f = \mathbf{1}_{KgK}$ on $\mathcal{Q}_K$ is given by $\varphi \mapsto q_{0!}p_0^*\varphi$, where $q_{0!}$ means summing over the fibers of q_0. Upon ignoring convergence issues, the trace of $R(\mathbf{1}_{KgK})$ on $\mathcal{Q}_K$ is equal to the cardinality of the restriction of $G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1})$ to the diagonal $G(F)\backslash G(\mathbb{A})/K$ via the maps (p_0, q_0). In other words, we should form the pullback diagram of groupoids

$$\begin{array}{ccc} M_{G,KgK} & \longrightarrow & G(F)\backslash G(\mathbb{A})/(K \cap gKg^{-1}) \\ \downarrow & & \downarrow (p_0,q_0) \\ G(F)\backslash G(\mathbb{A})/K & \xrightarrow{\Delta} & G(F)\backslash G(\mathbb{A})/K \times G(F)\backslash G(\mathbb{A})/K \end{array} \tag{2-3}$$

and we have a heuristic identity

$$\mathrm{Tr}(R(\mathbf{1}_{KgK}), \mathcal{Q}_K)'' ='' \#M_{G,KgK}. \tag{2-4}$$

Here $\#\mathcal{X}$ of a groupoid $\mathcal{X}$ is a counting of isomorphism classes of objects in $\mathcal{X}$ weighted by the reciprocals of the sizes of automorphism groups.

2.1.2 Weil's interpretation. Let $K_0 = \prod_{x\in|X|} G(\mathcal{O}_x)$. It was observed by Weil that the double coset groupoid $G(F)\backslash G(\mathbb{A})/K_0$ is naturally isomorphic to the groupoid of G-bundles over X that are trivial at the generic point of X. In fact, starting from $g = (g_x) \in G(\mathbb{A})$, one assigns the G-bundle on X that is glued from the trivial bundles on the generic point $\mathrm{Spec}\, F$ and the formal disks $\mathrm{Spec}\, \mathcal{O}_x$ via the "transition matrices" g_x. For a compact open $K \subset K_0$, one can similarly interpret $G(F)\backslash G(\mathbb{A})/K$ as the groupoid of G-bundles with K-level structures. There is an algebraic stack $\mathrm{Bun}_{G,K}$ classifying G-bundles on X with K-level structures, and the above observation can be rephrased as a fully faithful embedding of groupoids

$$G(F)\backslash G(\mathbb{A})/K \hookrightarrow \mathrm{Bun}_{G,K}(k). \tag{2-5}$$

A priori, the groupoid $\mathrm{Bun}_{G,K}(k)$ contains also G-bundles that are not trivial at the generic point, or equivalently G'-bundles for certain inner forms G' of G. Since we assume G is split, the embedding (2-5) is in fact an equivalence.

In the same spirit, we interpret $G(F)\backslash G(\mathbb{A})/(K\cap gKg^{-1})$ as the groupoid of triples $(\mathcal{E},\mathcal{E}',\alpha)$ where $\mathcal{E},\mathcal{E}'$ are G-bundles with K-level structures on X, and $\alpha:\mathcal{E}\dashrightarrow\mathcal{E}'$ is a *rational* isomorphism between $\mathcal{E}$ and $\mathcal{E}'$ (i.e., an isomorphism of G-bundles over the generic point Spec F) such that the relative position of $\mathcal{E}$ and $\mathcal{E}'$ at each closed point $x\in|X|$ is given by the double coset $K_xg_xK_x$. For example, when $G=\mathrm{GL}_n$, $K_x=\mathrm{GL}_n(\mathcal{O}_x)$ and $g_x=\mathrm{diag}(\varpi_x,1,\cdots,1)$, then α has relative position $K_xg_xK_x$ at x if and only if α extends to a homomorphism $\alpha_x:\mathcal{E}|_{\mathrm{Spec}\,\mathcal{O}_x}\to\mathcal{E}'|_{\mathrm{Spec}\,\mathcal{O}_x}$, and that $\mathrm{coker}(\alpha_x)$ is one-dimensional over the residue field k_x. There is a moduli stack $\mathrm{Hk}_{G,KgK}$ classifying such triples $(\mathcal{E},\mathcal{E}',\alpha)$. The above discussion can be rephrased as an equivalence of groupoids

(2-6) $$\mathrm{Hk}_{G,KgK}(k)\cong G(F)\backslash G(\mathbb{A})/(K\cap gKg^{-1}).$$

Moreover, $\mathrm{Hk}_{G,KgK}$ is equipped with two maps to $\mathrm{Bun}_{G,K}$ by recording $\mathcal{E}$ and $\mathcal{E}'$, which allow us to view it as a self-correspondence of $\mathrm{Bun}_{G,K}$

(2-7) $$\mathrm{Bun}_{G,K}\xleftarrow{p}\mathrm{Hk}_{G,KgK}\xrightarrow{q}\mathrm{Bun}_{G,K}$$

Under the equivalences (2-5) and (2-6), the diagram (2-7) becomes the diagram (2-2) after taking k-points.

2.1.3 Geometric interpretation of the trace. Continuing further with Weil's observation, we can form the stack-theoretic version of (2-3), and define the stack $\mathfrak{M}_{G,KgK}$ by the Cartesian diagram

(2-8) $$\begin{array}{ccc}\mathfrak{M}_{G,KgK} & \longrightarrow & \mathrm{Hk}_{G,KgK}\\ \downarrow & & \downarrow{\scriptstyle(p,q)}\\ \mathrm{Bun}_{G,K} & \xrightarrow{\Delta} & \mathrm{Bun}_{G,K}\times\mathrm{Bun}_{G,K}\end{array}$$

so that $\mathfrak{M}_{G,KgK}(k)=M_{G,KgK}$.

By the defining Cartesian diagram (2-8), $\mathfrak{M}_{G,KgK}$ classifies pairs $(\mathcal{E},\varphi)$ where $\mathcal{E}$ is a G-bundle over X with K-level structures, and $\varphi:\mathcal{E}\dashrightarrow\mathcal{E}$ is a rational *automorphism* with relative position given by KgK.

Recall the classical Hitchin moduli stack $\mathfrak{M}_{G,\mathcal{L}}$ in Section 1.1. If we write $\mathcal{L}=\mathcal{O}_X(D)$ for some effective divisor D, $\mathfrak{M}_{G,\mathcal{L}}$ then classifies pairs $(\mathcal{E},\varphi)$ where $\mathcal{E}$ is a G-bundle over X and φ is a rational section of $\mathrm{Ad}(\mathcal{E})$ (an infinitesimal automorphism of $\mathcal{E}$) with poles controlled by D. Therefore $\mathfrak{M}_{G,\mathcal{L}}$ can be viewed as a Lie algebra version of $\mathfrak{M}_{G,KgK}$, and $\mathfrak{M}_{G,KgK}$ is a group version of $\mathfrak{M}_{G,\mathcal{L}}$.

Let $\mathfrak{C}_G$ be the GIT quotient of G by the conjugation action of G; $\mathfrak{C}_G(F)$ is the set of *stable conjugacy classes* in $G(F)$. There is an affine scheme $\mathfrak{B}_{G,KgK}$ classifying rational maps $X\dashrightarrow\mathfrak{C}_G$ with poles controlled by KgK. The Hitchin fibration (1-1) has an analogue

$$h_G : \mathfrak{M}_{G,KgK} \to \mathfrak{B}_{G,KgK}. \tag{2-9}$$

Using the map h_G, the counting of $M_{G,KgK}$, hence the trace of $R(\mathbf{1}_{KgK})$, can be decomposed into a sum over certain stable conjugacy classes a

$$\mathrm{Tr}(R(\mathbf{1}_{KgK}), \mathfrak{A}_K)'' ='' \sum_{a\in\mathfrak{B}_{G,KgK}(k)} \#\mathfrak{M}_{G,KgK}(a)(k). \tag{2-10}$$

Here $\mathfrak{M}_{G,KgK}(a)$ (a stack over k) is the fiber $h_G^{-1}(a)$. To tie back to the classical story, $\#\mathfrak{M}_{G,KgK}(a)(k)$ is in fact a sum of orbital integrals

$$\#\mathfrak{M}_{G,KgK}(a)(k) = \sum_{\gamma\in G(F)/\sim,[\gamma]=a} J_\gamma(\mathbf{1}_{KgK})$$

over $G(F)$-conjugacy classes γ that belong to the stable conjugacy class a.

By the Lefschetz trace formula, we can rewrite (2-10) as

$$\mathrm{Tr}(R(\mathbf{1}_{KgK}), \mathfrak{A}_K)'' ='' \sum_{a\in\mathfrak{B}_{G,KgK}(k)} \mathrm{Tr}(\mathrm{Frob}_a, (\mathbf{R}h_{G!}\mathbb{Q}_\ell)_a). \tag{2-11}$$

This formula relates the Arthur-Selberg trace to the direct image complex of the Hitchin fibration h_G (called the *Hitchin complex* for G). Although it is still difficult to get a closed formula for each term in (2-11), this geometric point of view can be powerful in comparing traces for two different groups G and H by relating their Hitchin bases and Hitchin complexes.

In the work of Ngô [2010], where the Lie algebra version was considered, the Hitchin complex was studied in depth using tools such as perverse sheaves and the decomposition theorem. When H is an endoscopic group of G, Ngô shows that the stable part of the Hitchin complex for H appears as a direct summand of the Hitchin complex for G, from which he deduces the Langlands-Shelstad fundamental lemma for Lie algebras over function fields.

2.2 Relative trace formulae.

2.2.1 Periods of automorphic forms. For simplicity we assume G is semisimple. Let $H \subset G$ be a subgroup defined over F, and μ_H a Haar measure on $H(\mathbb{A})$. For a cuspidal automorphic representation π of $G(\mathbb{A})$, the linear functional on π

$$\begin{aligned} \mathcal{P}^G_{H,\pi} : \pi &\to \mathbb{C} \\ \varphi &\mapsto \int_{H(F)\backslash H(\mathbb{A})} \varphi\mu_H \end{aligned} \tag{2-12}$$

is called the H-*period* of π. It factors through the space of coinvariants $\pi_{H(\mathbb{A})}$.

One can also consider variants where we integrate φ against an automorphic character χ of $H(\mathbb{A})$. If π has nonzero H-period, it is called H-distinguished. Distinguished representations are used to characterize important classes of automorphic representations such as those coming by functoriality from another group. In case the local coinvariants $(\pi_x)_{H(F_x)}$ are one-dimensional for almost all places x (as in the case for many spherical subgroups of G), one expects the period $\mathcal{P}^G_{H,\pi}$ to be related to special values of L-functions of π.

2.2.2 Example. Let $G = \mathrm{PGL}_2$ and $H = A$ be the diagonal torus. Then by Hecke's theory, for a suitably chosen $\varphi \in \pi$, $\int_{H(F)\backslash H(\mathbb{A})} \varphi(t)|t|^{s-1/2}dt$ gives the standard L-function $L(s,\pi)$.

2.2.3 Relative trace formulae. Now suppose H_1, H_2 are two subgroups of G. Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ and $\widetilde{\pi}$ its contragradient. We get a bilinear form

$$\mathcal{P}^G_{H_1,\pi} \otimes \mathcal{P}^G_{H_2,\widetilde{\pi}} : \pi \otimes \widetilde{\pi} \to \mathbb{C}$$

In case the local coinvariants $(\pi_x)_{H_i(F_x)}$ are one-dimensional for all places x and $i = 1, 2$, the $H_1(\mathbb{A}) \times H_2(\mathbb{A})$ invariant bilinear forms on $\pi \times \widetilde{\pi}$ are unique up to scalar. Therefore $\mathcal{P}^G_{H_1,\pi} \otimes \mathcal{P}^G_{H_2,\widetilde{\pi}}$ is a multiple of the natural pairing on $\pi \times \widetilde{\pi}$ given by the Petersson inner product. This multiple is often related to special values of L-functions attached to π. For a systematic treatment of this topic, see Sakellaridis and Venkatesh [n.d.].

An important tool to study the pairing $\mathcal{P}^G_{H_1,\pi} \otimes \mathcal{P}^G_{H_2,\widetilde{\pi}}$ is the relative trace formula. We have natural maps of cosets

(2-13) $$H_1(F)\backslash H_1(\mathbb{A}) \xrightarrow{\varphi_1} G(F)\backslash G(\mathbb{A}) \xleftarrow{\varphi_2} H_2(F)\backslash H_2(\mathbb{A})$$

Consider the push-forward of the constant functions on $H_i(F)\backslash H_i(\mathbb{A})$ along φ_i, viewed as distributions on $G(F)\backslash G(\mathbb{A})$. Since we will only give a heuristic discussion of the relative trace formula, we will pretend that the L^2-pairing of two distributions makes sense. The relative trace of a test function $f \in C_c^\infty(G(\mathbb{A}))$ is the L^2-pairing

(2-14) $$\mathrm{RTr}^G_{H_1,H_2}(f) \text{''}=\text{''} \langle \varphi_{1,!}\mathbf{1}_{H_1(F)\backslash H_1(\mathbb{A})}, R(f)\varphi_{2,!}\mathbf{1}_{H_2(F)\backslash H_2(\mathbb{A})}\rangle_{L^2(G(F)\backslash G(\mathbb{A}),\mu_G)}$$

A variant of this construction is to replace the constant function 1 on $H_i(F)\backslash H_i(\mathbb{A})$ by an automorphic quasi-character χ_i of $H_i(\mathbb{A})$.

The relative trace formula is an equality between a spectral expansion of $\mathrm{RTr}^G_{H_1,H_2}(f)$ into quantities related to $\mathcal{P}^G_{H_1,\pi} \otimes \mathcal{P}^G_{H_2,\widetilde{\pi}}$ and a geometric expansion into a sum of orbital integrals $J^G_{H_1,H_2,\gamma}(f)$ indexed by double cosets $H_1(F)\gamma H_2(F) \subset G(F)$

(2-15)
$$J^G_{H_1,H_2,\gamma}(f) = \mathrm{vol}(H_\gamma(F)\backslash H_\gamma(\mathbb{A}), \mu_{H_\gamma}) \int_{H_\gamma(\mathbb{A})\backslash(H_1\times H_2)(\mathbb{A})} f(h_1^{-1}\gamma h_2)\frac{\mu_{H_1\times H_2}}{\mu_{H_\gamma}}(h_1,h_2)$$

where H_γ is the stabilizer of γ under the left-right translation on G by $H_1 \times H_2$.

Let $\mathfrak{C}^G_{H_1,H_2} = \mathrm{Spec}\ F[G]^{H_1\times H_2}$. We may call elements in $\mathfrak{C}^G_{H_1,H_2}(F)$ *stable orbits* of G under the action of $H_1 \times H_2$. There is a tautological map

$$\mathrm{inv} : H_1(F)\backslash G(F)/H_2(F) \to \mathfrak{C}^G_{H_1,H_2}(F).$$

We define

$$J^G_{H_1,H_2}(a, f) = \sum_{\gamma\in H_1(F)\backslash G(F)/H_2(F),\mathrm{inv}(\gamma)=a} J^G_{H_1,H_2,\gamma}(f). \tag{2-16}$$

Now we fix a compact open subgroup $K \subset G(\mathbb{A})$ and let $K_i = K\cap H_i(\mathbb{A})$ for $i = 1, 2$. Choose Haar measures on $H_i(\mathbb{A})$ and $G(\mathbb{A})$ so that K_i and K have volume 1. Consider the test function $f = \mathbf{1}_{KgK}$ as before. Unwinding the definition of the relative trace, we can rewrite (2-14) as

$$\mathrm{RTr}^G_{H_1,H_2}(\mathbf{1}_{KgK})'' ='' \#M^G_{H_1,H_2,KgK} \tag{2-17}$$

where $M^G_{H_1,H_2,KgK}$ is defined by the Cartesian diagram of groupoids

$$\begin{array}{ccc} M^G_{H_1,H_2,KgK} & \longrightarrow & G(F)\backslash G(\mathbb{A})/(K\cap gKg^{-1}) \\ \downarrow & & \downarrow{\scriptstyle (p_0,q_0)} \\ H_1(F)\backslash H_1(\mathbb{A})/K_1 \times H_2(F)\backslash H_2(\mathbb{A})/K_2 & \xrightarrow{(\varphi_1,\varphi_2)} & G(F)\backslash G(\mathbb{A})/K \times G(F)\backslash G(\mathbb{A})/K \end{array}$$

2.2.4 Geometric interpretation of the relative trace. Now we give a geometric interpretation of $M^G_{H_1,H_2,KgK}$. We assume H_1 and H_2 are also obtained by base change from split reductive groups over k, which we denote by the same notation. We have maps

$$\mathrm{Bun}_{H_1,K_1} \xrightarrow{\Phi_1} \mathrm{Bun}_{G,K} \xleftarrow{\Phi_2} \mathrm{Bun}_{H_2,K_2}.$$

Taking k-points of the above diagram we recover (2-13) up to modding out by the compact open subgroups K_i and K. We may form the Cartesian diagram of stacks

$$\begin{array}{ccc} \mathfrak{M}^G_{H_1,H_2,KgK} & \longrightarrow & \mathrm{Hk}_{G,KgK} \\ \downarrow & & \downarrow{\scriptstyle (p,q)} \\ \mathrm{Bun}_{H_1,K_1} \times \mathrm{Bun}_{H_2,K_2} & \xrightarrow{(\Phi_1,\Phi_2)} & \mathrm{Bun}_{G,K} \times \mathrm{Bun}_{G,K} \end{array}$$

We have $\mathfrak{M}^G_{H_1,H_2,KgK}(k) = M^G_{H_1,H_2,KgK}$.

2.2.5 Example. Consider the special case $G = G_1 \times G_1$ (G_1 is a semisimple group over k) and $H_1 = H_2$ is the diagonal copy of G_1, $K_1 = K_2$, $K = K_1 \times K_2$. Taking $g = (1, g_1)$ for some $g_1 \in G_1(\mathbb{A})$, we get a canonical isomorphism between the stacks $\mathfrak{M}^{G_1\times G_1}_{\Delta(G_1),\Delta(G_1),K(1,g_1)K}$ and $\mathfrak{M}_{G_1,K_1g_1K_1}$. In this case, the relative trace is the usual trace of $R(\mathbf{1}_{K_1g_1K_1})$ on the space of automorphic forms for G_1.

The moduli stack $\mathfrak{M}^G_{H_1,H_2,KgK}$ classifies $(\mathcal{E}_1, \mathcal{E}_2, \alpha)$ where $\mathcal{E}_i$ is an H_i-bundle with K_i-structure over X for $i = 1, 2$; α is a rational isomorphism between the G-bundles induced from $\mathcal{E}_1$ and $\mathcal{E}_2$, with relative position given by KgK.

One can construct a scheme $\mathfrak{B}^G_{H_1,H_2,KgK}$ classifying rational maps $X \dashrightarrow \mathfrak{C}^G_{H_1,H_2}$ with poles controlled by KgK, so that $\mathfrak{B}^G_{H_1,H_2,KgK}(k) \subset \mathfrak{C}^G_{H_1,H_2}(F)$. For $(\mathcal{E}_1, \mathcal{E}_2, \alpha) \in \mathfrak{M}^G_{H_1,H_2,KgK}$, we may restrict α to the generic point of X and take its invariants as a rational map $X \dashrightarrow \mathfrak{C}^G_{H_1,H_2}$. This way we get a map of algebraic stacks

$$h^G_{H_1,H_2} : \mathfrak{M}^G_{H_1,H_2,KgK} \to \mathfrak{B}^G_{H_1,H_2,KgK}. \tag{2-18}$$

In the situation of Example 2.2.5, $h^G_{H_1,H_2}$ specializes to the usual Hitchin map (2-9) for G_1, so we may think of $h^G_{H_1,H_2}$ as an analogue of the Hitchin map for the Hitchin-like moduli $\mathfrak{M}^G_{H_1,H_2,KgK}$. Taking k-points of (2-18) we get a map

$$M^G_{H_1,H_2,KgK} \to [H_1\backslash G/H_2](F) \xrightarrow{\text{inv}} \mathfrak{C}^G_{H_1,H_2}(F) \tag{2-19}$$

whose fiber over $a \in \mathfrak{C}^G_{H_1,H_2}(F)$ has cardinality equal to $J^G_{H_1,H_2}(a, \mathbf{1}_{KgK})$ defined in (2-16). We may thus decompose the relative trace into a sum of point-counting along the fibers of the map (2-19)

$$\begin{aligned} \mathrm{RTr}^G_{H_1,H_2}(\mathbf{1}_{KgK})'' =''& \sum_{a\in\mathfrak{C}^G_{H_1,H_2}(F)} J^G_{H_1,H_2}(a, \mathbf{1}_{KgK}) \\ =& \sum_{a\in\mathfrak{B}^G_{H_1,H_2,KgK}(k)} \mathrm{Tr}(\mathrm{Frob}_a, (\mathbf{R}h^G_{H_1,H_2,!}\mathbb{Q}_\ell)_a). \end{aligned} \tag{2-20}$$

The above formula relates the relative trace to the direct image complex $\mathbf{R}h^G_{H_1,H_2,!}\mathbb{Q}_\ell$. As in the case of the Arthur-Selberg trace formula, we may apply sheaf-theoretic tools to study this direct image complex, especially when it comes to comparing two such complexes.

2.2.6 Example. Consider the case $G = \mathrm{PGL}_2$, and $H_1 = H_2 = A \subset G$ is the diagonal torus. Let $K_1 = K_2 = \prod_x A(\mathcal{O}_x)$ and $K = \prod_x G(\mathcal{O}_x)$. Let $D = \sum_x n_x x$ be an effective divisor. Define the function h_D on $G(\mathbb{A})$ to be the characteristic function of $\mathrm{Mat}_2(\mathbb{O})_D = \{(g_x) | g_x \in \mathrm{Mat}_2(\mathcal{O}_x), v_x(\det g_x) = n_x, \forall x \in |X|\}$. Then $\mathrm{Mat}_2(\mathbb{O})_D$ is a finite union of K-double cosets Kg_iK. We will define a stack $\mathfrak{M}^G_{A,A,D}$ which turns out to be the union of the $\mathfrak{M}^G_{A,A,Kg_iK}$.

Consider the stack $\widetilde{\mathfrak{M}}^G_{A,A,D}$ classifying the data $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2, \varphi)$ where

- $\mathcal{L}_i$, $\mathcal{L}'_i$ are line bundles over X, for $i = 1, 2$;
- $\varphi : \mathcal{L}_1 \oplus \mathcal{L}_2 \to \mathcal{L}'_1 \oplus \mathcal{L}'_2$ is an injective map of coherent sheaves such that $\det(\varphi)$, viewed as a section of the line bundle $\mathcal{L}_1^{-1} \otimes \mathcal{L}_2^{-1} \otimes \mathcal{L}'_1 \otimes \mathcal{L}'_2$, has divisor D.

The Picard stack Pic_X acts on $\widetilde{\mathfrak{M}}^G_{A,A,D}$ by simultaneously tensoring on $\mathcal{L}_i$ and $\mathcal{L}'_i$. We define

$$\mathfrak{M}^G_{A,A,D} \cong \widetilde{\mathfrak{M}}^G_{A,A,D} / \mathrm{Pic}_X.$$

The bi-A-invariant regular functions on G are generated by $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \frac{bc}{ad-bc}$, therefore the space $\mathfrak{C}^G_{A,A}$ is isomorphic to $\mathbb{A}^1$. Define the Hitchin base $\mathfrak{B}^G_{A,A,D}$ to be the affine space $\mathrm{H}^0(X, \mathcal{O}_X(D))$.

To define the Hitchin map in this case, we write φ above as a matrix $\begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix}$ where φ_{ij} is a section of $\mathcal{L}_j^{-1} \otimes \mathcal{L}'_i$. The determinant $\det(\varphi)$ gives an isomorphism $\mathcal{L}_1^{-1} \otimes \mathcal{L}_2^{-1} \otimes \mathcal{L}'_1 \otimes \mathcal{L}'_2 \cong \mathcal{O}_X(D)$. On the other hand, $\varphi_{12}\varphi_{21}$ gives another section of $\mathcal{L}_1^{-1} \otimes \mathcal{L}_2^{-1} \otimes \mathcal{L}'_1 \otimes \mathcal{L}'_2$. The Hitchin map

$$h^G_{A,A} : \mathfrak{M}^G_{A,A,D} \to \mathrm{H}^0(X, \mathcal{O}_X(D)) = \mathfrak{B}^G_{A,A,D}$$

then sends $(\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2, \varphi)$ to $\varphi_{12}\varphi_{21}$, viewed as a section of $\mathcal{O}_X(D)$ via the identification given by $\det(\varphi)$.

Although $\mathfrak{M}^G_{A,A,D}$ is not of finite type, it is the disjoint union of finite type substacks indexed by a subset of $\mathbb{Z}^4/\Delta(\mathbb{Z})$. Indeed, for $\underline{d} = (d_1, d_2, d'_1, d'_2) \in \mathbb{Z}^4/\Delta(\mathbb{Z})$ such that $d'_1 + d'_2 = d_1 + d_2 + \deg D$, the substack ${}^{\underline{d}}\mathfrak{M}^G_{A,A,D}$ where $\deg \mathcal{L}_i = d_i$ and $\deg \mathcal{L}'_i = d'_i$ is of finite type. We may write $\mathrm{RTr}^G_{A,A}(h_D)$ as a formal sum of

$${}^{\underline{d}}\mathrm{RTr}^G_{A,A}(h_D) = \#{}^{\underline{d}}\mathfrak{M}^G_{A,A,D}(k) = \mathrm{Tr}(\mathrm{Frob}, \mathrm{H}^*_c({}^{\underline{d}}\mathfrak{M}^G_{A,A,D} \otimes \overline{k}, \overline{\mathbb{Q}}_\ell)).$$

2.3 Relative fundamental lemma. In many cases we do not expect to prove closed formulae for relative traces of the form (2-14). Instead, for applications to problems on automorphic representations, it often suffices to establish an identity between relative traces for two different situations (G, H_1, H_2) and (G', H'_1, H'_2).

2.3.1 General format of RTF comparison. In order to establish such an identity, we need the following structures or results:

1. There should be an isomorphism between the spaces of invariants $\mathfrak{C}^G_{H_1,H_2} \cong \mathfrak{C}^{G'}_{H'_1,H'_2}$.

2. (fundamental lemma) For almost all $x \in |X|$, and

$$a_x \in \mathfrak{C}^G_{H_1,H_2}(F_x) \cong \mathfrak{C}^{G'}_{H'_1,H'_2}(F_x)$$

   we should have an identity of local orbital integrals up to a transfer factor

$$J^G_{H_1,H_2,x}(a_x, \mathbf{1}_{G(\mathcal{O}_x)}) \sim J^{G'}_{H'_1,H'_2,x}(a_x, \mathbf{1}_{G'(\mathcal{O}_x)}).$$

   Here $J^G_{H_1,H_2,x}(a_x, f_x)$ is the local analogue of $J^G_{H_1,H_2}(a, f)$ defined in (2-16).

3. (smooth matching) For any $x \in |X|$, $a_x \in \mathfrak{C}^G_{H_1,H_2}(F_x) \cong \mathfrak{C}^{G'}_{H'_1,H'_2}(F_x)$, and $f_x \in C^\infty_c(G(F_x))$, there exists $f'_x \in C^\infty_c(G'(F_x))$ such that $J^G_{H_1,H_2,x}(a_x, f_x) = J^{G'}_{H'_1,H'_2,x}(a_x, f'_x)$.

The geometric interpretation (2-20) of the relative trace gives a way to prove the fundamental lemma by comparing the direct image complexes of the Hitchin maps $h^G_{H_1,H_2}$ and $h^{G'}_{H'_1,H'_2}$. Below we discuss one such example.

2.3.2 The relative trace formulae of Jacquet and Rallis. 2011 proposed a relative trace formula approach to the Gan-Gross-Prasad conjecture for unitary groups. They considered two relative trace formulae, one involving general linear groups, and the other involving unitary groups. They formulated both the fundamental lemma and the smooth matching in this context as conjectures. In Yun [2011b], we used the geometric interpretation sketched in Section 2.2.4 to prove the fundamental lemma conjectured by Jacquet and Rallis, in the case of function fields. In the appendix to Yun [ibid.], J.Gordon used model theory to deduce the mixed characteristic case from the function field case. On the other hand, W. Zhang [2014] proved the smooth matching for the Jacquet-Rallis relative trace formula at non-archimedean places. Together with the fundamental lemma proved in Yun [2011b], W.Zhang deduced the Gan-Gross-Prasad conjecture for unitary groups under some local restrictions.

In the next two examples, we introduce the groups involved in the two trace formulae in Jacquet and Rallis [2011], and sketch the definition of the moduli stacks relevant to the orbital integrals. Since we proved the fundamental lemma by reducing to its Lie algebra analogue, our moduli stacks will be linearized versions of the Hitchin-type moduli stacks introduced in Section 2.2.4, which are closer to the classical Hitchin moduli stack.

2.3.3 Example. Let F'/F be a separable quadratic extension corresponding to a double cover $\nu : X' \to X$. Let $\sigma \in \mathrm{Gal}(F'/F)$ be the nontrivial involution. Consider $G = \mathrm{Res}_{F'/F}\mathrm{GL}_n \times \mathrm{Res}_{F'/F}\mathrm{GL}_{n-1}$, $H_1 = \mathrm{Res}_{F'/F}\mathrm{GL}_{n-1}$ and $H_2 = \mathrm{GL}_n \times \mathrm{GL}_{n-1}$ (over F). The embedding $H_1 \to G$ sends $h \in H_1$ to $(\begin{pmatrix} h & \\ & 1 \end{pmatrix}, h) \in G$.

The double quotient $H_1 \backslash G / H_2$ can be identified with $\mathrm{GL}_{n-1} \backslash \mathrm{S}_n$, where

$$\mathrm{S}_n = \{g \in \mathrm{Res}_{F'/F}\mathrm{GL}_n | \sigma(g) = g^{-1}\}$$

with GL_{n-1} acting by conjugation. The local orbital integral relevant to this relative trace formula is

(2-21)
$$J^{\mathrm{GL}}_{x,\gamma}(f) := \int_{\mathrm{GL}_{n-1}(F_x)} f(h^{-1}\gamma h)\eta_x(\det h)dh, \quad \gamma \in \mathrm{S}_n(F_x), f \in C_c^\infty(\mathrm{S}_n(F_x)).$$

Here η_x is the character $F_x^\times \to \{\pm 1\}$ attached to the quadratic extension F'_x/F_x.

The Lie algebra analogue of $\mathrm{GL}_{n-1}\backslash \mathrm{S}_n$ is $\mathrm{GL}_{n-1}\backslash(\mathfrak{gl}_n \otimes_F F'_-)$ where $F'_- = (F')^{\sigma=-1}$, and GL_{n-1} acts by conjugation. Let V_n be the standard representation of GL_n over F. It is more convenient to identify $\mathrm{GL}_{n-1}\backslash(\mathfrak{gl}_n \otimes_F F'_-)$ with

$$\mathrm{GL}_n \backslash \left(\mathrm{Hom}_F(V_n, V_n \otimes F'_-) \times (V_n \times V_n^*)^1\right),$$

where $(V_n \times V_n^*)^1$ consists of $(e, e^*) \in V_n \times V_n^*$ such that $e^*(e) = 1$, and GL_n is acting diagonally on all factors (conjugation on the first factor). The GIT quotient $\mathfrak{C}$ of $\mathrm{Hom}_F(V_n, V_n \otimes F'_-) \times (V_n \times V_n^*)^1$ by GL_n is an affine space of dimension $2n-1$. For $(\varphi, e, e^*) \in \mathrm{Hom}_F(V_n, V_n \otimes F'_-) \times (V_n \times V_n^*)^1$, we have invariants $a_i(\varphi) \in (F'_-)^{\otimes i}$ that records the i-th coefficient of the characteristic polynomial of φ ($1 \le i \le n$), and $b_i = e^*(\varphi^i e) \in (F'_-)^{\otimes i}$ for $1 \le i \le n-1$. The invariants $(a_1, \cdots, a_n, b_1, \cdots, b_{n-1})$ give coordinates for $\mathfrak{C}$.

We introduce the following moduli stack $\mathfrak{M}$ which serves as a global avatar for the Lie algebra version of the orbital integrals appearing in this relative trace formula. Fix line bundles $\mathfrak{L}$ and $\mathfrak{L}'$ on X. Let $\mathfrak{L}_- = \mathfrak{L} \otimes_{\mathcal{O}_X} \mathcal{O}_{X'}^{\sigma=-1}$. The stack $\mathfrak{M}$ classifies tuples $(\mathcal{E}, \varphi, s, s^*)$ where $\mathcal{E}$ is a vector bundle of rank n over X, $\varphi : \mathcal{E} \to \mathcal{E} \otimes \mathfrak{L}_-$, $s : \mathfrak{L}'^{-1} \to \mathcal{E}$ and $s^* : \mathcal{E} \to \mathfrak{L}'$ are $\mathcal{O}_X$-linear maps of coherent sheaves. The "Hitchin base" $\mathfrak{B}$ in this situation is the affine space $\prod_{i=1}^n \Gamma(X, \mathfrak{L}_-^{\otimes i}) \times \prod_{i=0}^{n-1} \Gamma(X, \mathfrak{L}'^{\otimes 2} \otimes \mathfrak{L}_-^{\otimes i})$. The Hitchin map $f : \mathfrak{M} \to \mathfrak{B}$ sends $(\mathcal{E}, \varphi, s, s^*)$ to the point of $\mathfrak{B}$ with coordinates $(a_1(\varphi), \cdots, a_n(\varphi), b_0, \cdots, b_{n-1})$, where $a_i(\varphi)$ are the coefficients of the characteristic polynomial of φ, and $b_i = s^* \circ \varphi^i \circ s : \mathfrak{L}'^{-1} \to \mathfrak{L}_-^{\otimes i} \otimes \mathfrak{L}'$.

2.3.4 Example. Let F'/F and $\nu : X' \to X$ be as in Example 2.3.3. Let W_{n-1} be a Hermitian vector space of dimension $n-1$ over F'. Let $W_n = W_{n-1} \oplus F'e_n$ with the

Hermitian form $(\cdot,\cdot)$ extending that on W_{n-1} and such that $W_{n-1} \perp e_n$, $(e_n, e_n) = 1$. Let U_n and U_{n-1} be the unitary groups over F attached to W_n and W_{n-1}. Consider $G' = \mathrm{U}_n \times \mathrm{U}_{n-1}$, and the subgroup $H'_1 = H'_2 = \mathrm{U}_{n-1}$ diagonally embedded into G'.

The double quotient $H'_1 \backslash G' / H'_2$ can be identified with the quotient $\mathrm{U}_{n-1} \backslash \mathrm{U}_n$ where U_{n-1} acts by conjugation. For $x \in |X|$, the local orbital integral relevant to this relative trace formula is

$$J^{\mathrm{U}}_{x,\delta}(f) = \int_{\mathrm{U}_{n-1}(F_x)} f(h^{-1}\delta h)dh, \quad \delta \in \mathrm{U}_n(F_x), f \in C_c^\infty(\mathrm{U}_n(F_x)).$$

The Lie algebra analogue of $\mathrm{U}_{n-1} \backslash \mathrm{U}_n$ is $\mathrm{U}_{n-1} \backslash \mathfrak{u}_n$, where $\mathfrak{u}_n$, the Lie algebra of U_n, consists of skew-self-adjoint endomorphisms of W_n. As in the case of Example 2.3.3, we identify $\mathrm{U}_{n-1} \backslash \mathfrak{u}_n$ with $\mathrm{U}_n \backslash \left(\mathfrak{u}_n \times W_n^1\right)$ where W_n^1 is the set of vectors $e \in W_n$ such that $(e, e) = 1$. The GIT quotient of $\mathfrak{u}_n \times W_n^1$ by U_n can be identified with the space $\mathfrak{C}$ introduced in Example 2.3.3. For $(\psi, e) \in \mathfrak{u}_n \times W_n^1$, its image in $\mathfrak{C}$ is $(a_1(\psi), \cdots, a_n(\psi), b_1, \cdots, b_{n-1})$ where $a_i(\psi) \in (F'_-)^{\otimes i}$ are the coefficients of the characteristic polynomial of ψ (as an endomorphism of W_n), and $b_i = (\psi^i e, e) \in (F'_-)^{\otimes i}$, since $\sigma(\psi^i e, e) = (e, \psi^i e) = (-1)^i(\psi^i e, e)$.

We introduce a moduli stack $\mathfrak{N}$ which serves as a global avatar for the Lie algebra version of the orbital integrals appearing in this relative trace formula. Fix line bundles $\mathfrak{L}$ and $\mathfrak{L}'$ on X. The stack $\mathfrak{N}$ classifies tuples $(\mathfrak{F}, h, \psi, t)$ where $\mathfrak{F}$ is a vector bundle of rank n on X', $h : \mathfrak{F} \xrightarrow{\sim} \sigma^*\mathfrak{F}^\vee$ is a Hermitian form on $\mathfrak{F}$, $\psi : \mathfrak{F} \to \mathfrak{F} \otimes \nu^*\mathfrak{L}$ is skew-self-adjoint with respect to h and $t : \nu^*\mathfrak{L}'^{-1} \to \mathfrak{F}$ is an $\mathcal{O}_{X'}$-linear map. When ν is unramified, the base $\mathfrak{B}$ introduced in Example 2.3.3 still serves as the Hitchin base for $\mathfrak{N}$. The Hitchin map $g : \mathfrak{N} \to \mathfrak{B}$ sends $(\mathfrak{F}, h, \psi, t)$ to $(a_1(\psi), \cdots, a_n(\psi), b_0, \cdots, b_{n-1})$, where $a_i(\psi)$ are the coefficients of the characteristic polynomial of ψ, and $b_i = t^\vee \circ h \circ \psi^i \circ t$ descends to an $\mathcal{O}_X$-linear map $\mathfrak{L}'^{-1} \to \mathfrak{L}_-^{\otimes i} \otimes \mathfrak{L}'$.

2.3.5 Theorem (Yun [2011b]). *Let x be a place of F such that F'/F is unramified over x and the Hermitian space $W_{n,x}$ has a self-dual lattices $\Lambda_{n,x}$. Then for strongly regular semisimple elements $\gamma \in \mathrm{S}_n(F_x)$ and $\delta \in \mathrm{U}_n(F_x)$ with the same invariants in $\mathfrak{C}(F_x)$, we have*

$$J^{\mathrm{GL}}_{x,\gamma}(\mathbf{1}_{\mathrm{S}_n(\mathcal{O}_x)}) = \pm J^{\mathrm{U}}_{x,\delta}(\mathbf{1}_{\mathrm{U}(\Lambda_{n,x})})$$

for some sign depending on the invariants of γ.

The main geometric observation in Yun [ibid.] is that both $f : \mathfrak{M} \to \mathfrak{B}$ and $g : \mathfrak{N} \to \mathfrak{B}$ are small maps when restricted to a certain open subset of $\mathfrak{B}$. This enables us to prove an isomorphism between the direct images complexes of f and g by checking over the generic point of $\mathfrak{B}$. Such an isomorphism of sheaves, after passing to the Frobenius traces on stalks, implies the identity above, which was the fundamental lemma conjectured by Jacquet and Rallis.

3 Hitchin moduli stack and Shtukas

In this section we consider automorphic objects that arise as cohomology classes of moduli stacks of Shtukas, which are the function-field counterpart of Shimura varieties. These cohomology classes generalize the notion of automorphic forms. The periods and relative traces have their natural analogues in this more general setting. Hitchin-type moduli stacks continue to play an important role in the study of such relative trace formulae. We give a survey of our recent work Yun and Zhang [2017],Yun and Zhang [n.d.] on higher Waldspurger-Gross-Zagier formulae and Yun [n.d.] on the arithmetic fundamental lemma, which fit into the framework to be discussed in this section.

3.1 Moduli of Shtukas. In his seminal paper, Drinfeld [1974] introduced the moduli of elliptic modules as a function field analogue of modular curves. Later, Drinfeld [1987] defined more general geometric object called Shtukas, and used them to prove the Langlands conjecture for GL_2 over function fields. Since then it became clear that the moduli stack of Shtukas should play the role of Shimura varieties for function fields, and its cohomology should realize the Langlands correspondence for global function fields. This idea was realized for GL_n by L. Lafforgue [2002] who proved the full Langlands conjecture in this case. For an arbitrary reductive group G, V. Lafforgue [2012] proved the automorphic to Galois direction of the Langlands conjecture using moduli stacks of Shtukas.

3.1.1 The moduli of Shtukas. The general definition of G-Shtukas was given by Varshavsky [2004]. For simplicity of presentation we assume G is split. Again we fix an open subgroup $K \subset K_0$, and let $N \subset |X|$ be the finite set of places where $K_x \neq G(\mathcal{O}_x)$. Choosing a maximal split torus T and a Borel subgroup B containing T, we may therefore talk about dominant coweights of T with respect to B. Let $r \geq 0$ be an integer. Let $\mu = (\mu_1, \cdots, \mu_r)$ be a sequence of dominant coweights of T. Recall dominant coweights of T are in bijection with relative positions of two G-bundles over the formal disk with the same generic fiber.

Let $\mathrm{Hk}^{\mu}_{G,K}$ be the Hecke stack classifying points $x_1, \cdots, x_r \in X - N$ together with a diagram of the form

$$\mathcal{E}_0 \overset{f_1}{\dashrightarrow} \mathcal{E}_1 \overset{f_2}{\dashrightarrow} \cdots \overset{f_r}{\dashrightarrow} \mathcal{E}_r$$

where $\mathcal{E}_i$ are G-bundles over X with K-level structures, and $f_i : \mathcal{E}_{i-1}|_{X-x_i} \xrightarrow{\sim} \mathcal{E}_i|_{X-x_i}$ is an isomorphism compatible with the level structures whose relative position at x_i is in the closure of that given by μ_i.

A *G-Shtuka of type μ with level K* is the same data as those classified by $\mathrm{Hk}^{\mu}_{G,K}$, together with an isomorphism of G-bundles compatible with K-level structures

$$\iota : \mathcal{E}_r \xrightarrow{\sim} {}^{\tau}\mathcal{E}_0. \tag{3-1}$$

Here, ${}^{\tau}\mathcal{E}_0$ is the image of $\mathcal{E}_0$ under the Frobenius morphism $\mathrm{Fr} : \mathrm{Bun}_{G,K} \to \mathrm{Bun}_{G,K}$. If we are talking about an S-family of G-Shtukas for some k-scheme S, $\mathcal{E}_0$ is a G-torsor over $X \times S$, then ${}^{\tau}\mathcal{E}_0 := (\mathrm{id}_X \times \mathrm{Fr}_S)^*\mathcal{E}_0$. There is a moduli stack $\mathrm{Sht}^{\mu}_{G,K}$ of G-Shtuka of type μ, which fits into a Cartesian diagram

$$\begin{array}{ccc} \mathrm{Sht}^{\mu}_{G,K} & \longrightarrow & \mathrm{Hk}^{\mu}_{G,K} \\ \downarrow & & \downarrow{\scriptstyle (p_0,p_r)} \\ \mathrm{Bun}_{G,K} & \xrightarrow{(\mathrm{id},\mathrm{Fr})} & \mathrm{Bun}_{G,K} \times \mathrm{Bun}_{G,K} \end{array} \tag{3-2}$$

Let us observe the similarity with the definition of the (group version of) Hitchin stack in diagram (2-8): the main difference is that we are replacing the diagonal map of $\mathrm{Bun}_{G,K}$ by the graph of the Frobenius.

Recording only the points $x_1, \cdots, x_r$ gives a morphism

$$\pi^{\mu}_{G,K} : \mathrm{Sht}^{\mu}_{G,K} \to (X - N)^r$$

The datum μ is called *admissible* if $\sum_i \mu_i$ lies in the coroot lattice. The existence of an isomorphism (3-1) forces μ to be admissible. Therefore $\mathrm{Sht}^{\mu}_{G,K}$ is nonempty only when μ is admissible. When $r = 0$, $\mathrm{Sht}^{\mu}_{G,K}$ is the discrete stack given by the double coset $\mathrm{Bun}_{G,K}(k) = G(F)\backslash G(\mathbb{A})/K$. For μ admissible, we have

$$d_G(\mu) := \dim \mathrm{Sht}^{\mu}_{G,K} = \sum_i (\langle 2\rho_G, \mu_i\rangle + 1).$$

3.1.2 Hecke symmetry. Let $g \in G(\mathbb{A})$ and let S be the finite set of $x \in |X| - N$ such that $g_x \notin G(\mathcal{O}_x)$. There is a self-correspondence $\mathrm{Sht}^{\mu}_{G,KgK}$ (the dependence on g is only through the double coset KgK) of $\mathrm{Sht}^{\mu}_{G,K}|_{(X-N-S)^r}$ such that both maps to $\mathrm{Sht}^{\mu}_{G,K}|_{(X-N-S)^r}$ are finite étale. It then induces an endomorphism of the direct image complex $\mathbf{R}\pi^{\mu}_{G,K,!}\mathrm{IC}(\mathrm{Sht}^{\mu}_{G,K})|_{(X-N-S)^r}$. V. Lafforgue [2012] used his construction of excursion operators to extend this endomorphism to the whole complex $\mathbf{R}\pi^{\mu}_{G,K,!}\mathrm{IC}(\mathrm{Sht}^{\mu}_{G,K})$ over $(X - N)^r$. If we assign this endomorphism to the function $\mathbf{1}_{KgK}$, it extends by linearity to an action of the Hecke algebra $C_c(K\backslash G(\mathbb{A})/K)$ on the complex $\mathbf{R}\pi^{\mu}_{G,K,!}\mathrm{IC}(\mathrm{Sht}^{\mu}_{G,K})$, and hence on its geometric stalks and on its cohomology groups.

3.1.3 Intersection cohomology of $\mathrm{Sht}^{\mu}_{G,K}$. The singularities of the map $\pi^{\mu}_{G,K}$ are exactly the same as the product of the Schubert varieties $\mathrm{Gr}_{G,\leq\mu_i}$ in the affine Grassmannian Gr_G. It is expected that the complex $\mathbf{R}\pi^{\mu}_{G,K,!}\mathrm{IC}(\mathrm{Sht}^{\mu}_{G,K})$ should realize the global Langlands correspondence for G in a way similar to the Eichler-Shimura correspondence for modular curves. The phenomenon of endoscopy makes stating a precise conjecture quite subtle, but a rough form of the expectation is a $C_c(K\backslash G(\mathbb{A})/K)$-equivariant decomposition over $(X-N)^r$

$$\mathbf{R}\pi^{\mu}_{G,K,!}\mathrm{IC}(\mathrm{Sht}^{\mu}_{G,K})'' ='' \left(\bigoplus_{\pi \text{ cuspidal}} \pi^K \otimes (\boxtimes_{i=1}^{r}\rho_{\pi}^{\mu_i}[1])\right)\bigoplus(\text{Eisenstein part}). \tag{3-3}$$

Here π runs over cuspidal automorphic representations of $G(\mathbb{A})$ such that $\pi^K \neq 0$, ρ_π is the $\widehat{G}$-local system on $X-N$ attached to π by the Langlands correspondence, and $\rho_\pi^{\mu_i}$ is the local system obtained by the composition

$$\pi_1(X-N,*) \xrightarrow{\rho_\pi} \widehat{G}(\overline{\mathbb{Q}}_\ell) \to \mathrm{GL}(V(\mu_i))$$

where $V(\mu_i)$ is the irreducible representation of the dual group $\widehat{G}$ with highest weight μ_i.

One approach to prove (3-3) is to use trace formulae. One the one hand, consider the action of a Hecke operator composed with a power of Frobenius at some $x \in |X|-N$ acting on the geometric stalk at x of the left side of (3-3), which is $\mathrm{IH}^*_c(\mathrm{Sht}^{\mu}_{G,K,\overline{x}})$. The trace of this action can be calculated by the Lefschetz trace formula, and can be expressed as a sum of twisted orbital integrals. On the other hand, the trace of the same operator on the right side of (3-3) can be calculated by the Arthur-Selberg trace formula, and be expressed using orbital integrals. The identity (3-3) would then follow from an identity between the twisted orbital integrals and the usual orbital integrals that appear in both trace formulae, known as the base-change fundamental lemma.

The difficulty in implementing this strategy is that $\mathrm{Sht}^{\mu}_{G,K}$ is not of finite type, and both the Lefschetz trace and the Arthur-Selberg trace would be divergent. L. Lafforgue [2002] treated the case $G=\mathrm{GL}_n$ and $\mu=((1,0,\cdots,0),(0,\cdots,0,-1))$ by difficult analysis of the compactification of truncations of $\mathrm{Sht}^{\mu}_{G,K}$, generalizing the work of Drinfeld on GL_2.

3.1.4 Cohomological spectral decomposition. We discuss a weaker version of the spectral decomposition (3-3). As mentioned above, $\mathrm{Sht}^{\mu}_{G,K}$ is not of finite type, so its intersection cohomology is not necessarily finite-dimensional. One can present $\mathrm{Sht}^{\mu}_{G,K}$ as an increasing union of finite-type open substacks, but these substacks are not preserved by the Hecke correspondences. Despite all that, we expect nice finiteness properties of $\mathrm{IH}^*_c(\mathrm{Sht}^{\mu}_{G,K}\otimes\overline{k})$ as a Hecke module. More precisely, the spherical Hecke algebra $C_c(K_0^N\backslash G(\mathbb{A}^N)/K_0^N,\overline{\mathbb{Q}}_\ell)$ (superscript N means removing places in N) should act through a quotient $\overline{\mathcal{H}}^N$ (possibly depending on μ) which is a finitely generated algebra over $\overline{\mathbb{Q}}_\ell$, and that $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})$ should be a finitely generated module over $\overline{\mathcal{H}}^N$. Now viewing $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})$ as a coherent sheaf on $\overline{\mathcal{H}}^N$, we should get a canonical decomposition of it in terms of connected components of $\overline{\mathcal{H}}^N$. A coarser decomposition should take the following form

$$\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})=\bigoplus_{[P]}\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})_{[P]} \tag{3-4}$$

where $[P]$ runs over associated classes of parabolic subgroups of G. The support of $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})_{[P]}$ should be described using the analogous quotient of the Hecke algebra $\overline{\mathcal{H}}^N_L$ for the Levi factor L of P, via the Satake transform from the spherical Hecke algebra for G to the one for L. If G is semisimple, the part $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k},\overline{\mathbb{Q}}_\ell)_{[G]}$ should be finite-dimensional over $\overline{\mathbb{Q}}_\ell$.

In the simplest nontrivial case $G=\mathrm{PGL}_2$ we have proved the coarse decomposition.

3.1.5 Theorem (Yun and Zhang [2017],Yun and Zhang [n.d.]). *For $G=\mathrm{PGL}_2$, consider the moduli of Shtukas Sht^r_G without level structures of type $\mu=(\mu_1,\cdots,\mu_r)$ where each μ_i is the minuscule coweight. Then there is a decomposition of Hecke modules*

$$\mathrm{H}^{2r}_c(\mathrm{Sht}^r_G\otimes\overline{k})=(\oplus_\chi\mathrm{H}^{2r}_c(\mathrm{Sht}^r_G\otimes\overline{k})[\chi])\oplus\mathrm{H}^{2r}_c(\mathrm{Sht}^r_G\otimes\overline{k})_{\mathrm{Eis}}$$

where χ runs over a finite set of characters of the Hecke algebra $C_c(K_0\backslash G(\mathbb{A})/K_0)$, and the support of $\mathrm{H}^{2r}_c(\mathrm{Sht}^r_G\otimes\overline{k})_{\mathrm{Eis}}$ is defined by the Eisenstein ideal.

For $i\neq 2r$, $\mathrm{H}^i_c(\mathrm{Sht}^r_G\otimes\overline{k})$ is finite-dimensional.

Similar result holds for a version of Sht^r_G with Iwahori level structures.

We expect the similar techniques to work for general split G and general type μ.

Assume we have an analogue of the above theorem for G. Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ such that $\pi^K\neq 0$, then $C_c(K^N_0\backslash G(\mathbb{A}^N)/K^N_0,\overline{\mathbb{Q}}_\ell)$ acts on π^K by a character χ_π up to semisimplification. Suppose χ_π does not appear in the support of $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_G\otimes\overline{k})_{[P]}$ for any proper parabolic P (in which case we say χ_π is non-Eisenstein), then the generalized eigenspace $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})[\chi_\pi]$ is a finite-dimensional direct summand of $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})$ containing the contribution of π but possibly also companions of π with the same Hecke character χ_π away from N.

3.2 Heegner-Drinfeld cycles and periods. When $r=0$, the left side of (3-3) is simply the function space $C_c(G(F)\backslash G(\mathbb{A})/K)$ where cuspidal automorphic forms live. In general, we should think of cohomology classes in $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K,\overline{x}})$ or $\mathrm{IH}^*_c(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k})$

as generalizations of automorphic forms. We shall use this viewpoint to generalize some constructions in Section 2 from classical automorphic forms to cohomology classes of $\mathrm{Sht}^{\mu}_{G,K}$.

3.2.1 Heegner-Drinfeld cycles. Let $H \subset G$ be a subgroup defined over k with level group $K_H = K \cap H(\mathbb{A})$. It induces a map $\theta_{\mathrm{Bun}} : \mathrm{Bun}_{H,K_H} \to \mathrm{Bun}_{G,K}$.

Fix an integer $r \geq 0$. Let $\lambda = (\lambda_1, \cdots, \lambda_r)$ be an admissible sequence of dominant coweights of H; let $\mu = (\mu_1, \cdots, \mu_r)$ be an admissible sequence of dominant coweights of G. To relate $\mathrm{Sht}^{\lambda}_{H}$ to Sht^{μ}_{G}, we need to impose more restrictions on λ and μ. For two coweights μ, μ' of G we write $\mu \leq_G \mu'$ if for some (equivalently all) choices of a Borel $B' \subset G_{\overline{k}}$ and a maximal torus $T' \subset B'$, $\mu'_{B',T'} - \mu_{B',T'}$ is a sum of positive roots. Here $\mu_{B',T'}$ (resp. $\mu'_{B',T'}$) is the unique dominant coweight of T' conjugate to μ (resp. μ').

We assume that $\lambda_i \leq_G \mu_i$ for $0 \leq i \leq r$. In this case there is a natural morphism of Hecke stacks

$$\theta_{\mathrm{Hk}} : \mathrm{Hk}^{\lambda}_{H,K_H} \to \mathrm{Hk}^{\mu}_{G,K}$$

compatible with θ_{Bun}. The Cartesian diagram (3-2) and its counterpart for $\mathrm{Sht}^{\lambda}_{H}$ then induce a map over $(X-N)^r$

$$\theta : \mathrm{Sht}^{\lambda}_{H,K_H} \to \mathrm{Sht}^{\mu}_{G,K}.$$

If θ is proper, the image of the fundamental class of $\mathrm{Sht}^{\lambda}_{H,K_H}$ defines an algebraic cycle $\mathrm{Sht}^{\mu}_{G,K}$ which we call a *Heegner-Drinfeld cycle*.

3.2.2 Example. Consider the case $G = \mathrm{PGL}_2$, and $H = T$ is a non-split torus of the form $T = (\mathrm{Res}_{F'/F}\mathbb{G}_m)/\mathbb{G}_m$ for some quadratic extension F'/F. Since T is not a constant group scheme over X, our previous discussion does not directly apply, but we can easily define what a T-Shtuka is. The quadratic extension F' is the function field of a smooth projective curve X' with a degree two map $\nu : X' \to X$. Let $\lambda = (\lambda_1, \cdots, \lambda_r) \in \mathbb{Z}^r$ with $\sum_i \lambda_i = 0$, we may consider the moduli of rank one Shtukas $\mathrm{Sht}^{\lambda}_{\mathrm{GL}_1,X'}$ over X' of type λ. We define $\mathrm{Sht}^{\lambda}_{T}$ to be the quotient $\mathrm{Sht}^{\lambda}_{\mathrm{GL}_1,X'}/\mathrm{Pic}_X(k)$, where the discrete groupoid $\mathrm{Pic}_X(k)$ is acting by pulling back to X' and tensoring with rank one Shtukas. It can be shown that the projection $\mathrm{Sht}^{\lambda}_{T} \to X'^r$ is a finite étale Galois cover with Galois group $\mathrm{Pic}_{X'}(k)/\mathrm{Pic}_X(k)$. In particular, $\mathrm{Sht}^{\lambda}_{T}$ is a smooth and proper DM stack over k of dimension r.

Now let $\mu = (\mu_1, \cdots, \mu_r)$ be a sequence of dominant coweights of G. Then each μ_i can be identified with an element in $\mathbb{Z}_{\geq 0}$, with the positive coroot corresponding to 1. Admissibility of μ means that $\sum_i \mu_i$ is even. The condition $\lambda_i \leq_G \mu_i$ is saying that $|\lambda_i| \leq \mu_i$ and that $\mu_i - \lambda_i$ is even. When $\lambda_i \leq_G \mu_i$ for all i, the map $\theta : \mathrm{Sht}^{\lambda}_{T} \to$

Sht_G^μ simply takes a rank one Shtuka $(\{\mathcal{E}_i\};\{x'_i\})$ on X' and sends it to the direct image $(\{\nu_*\mathcal{E}_i\};\{\nu(x'_i)\})$, which is a rank two Shtuka on X.

3.2.3 Periods. Fix a Haar measure μ_H on $H(\mathbb{A})$. Under a purely root-theoretic condition on λ and μ, θ^* induces a map

$$\theta^* : \mathrm{IH}_c^{2d_H(\lambda)}(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k}) \to \mathrm{H}_c^{2d_H(\lambda)}(\mathrm{Sht}^\lambda_{H,K_H}\otimes\overline{k}),$$

and therefore defines a period map

$$\mathcal{P}^{G,\mu}_{H,\lambda} : \mathrm{IH}_c^{2d_H(\lambda)}(\mathrm{Sht}^\mu_{G,K}\otimes\overline{k}) \xrightarrow{\theta^*} \mathrm{H}_c^{2d_H(\lambda)}(\mathrm{Sht}^\lambda_{H,K_H}\otimes\overline{k}) \xrightarrow{\cap[\mathrm{Sht}^\lambda_{H,K_H}]\cdot\mathrm{vol}(K_H,\mu_H)} \overline{\mathbb{Q}}_\ell.$$

The last map above is the cap product with the fundamental class of Sht^λ_H followed by multiplication by $\mathrm{vol}(K_H,\mu_H)$.

Now assume $\mathrm{Sht}^\lambda_{H,K_H}$ has half the dimension of $\mathrm{Sht}^\mu_{G,K}$,

$$d_G(\mu) = 2d_H(\lambda). \tag{3-5}$$

Let π be a cuspidal automorphic representation of $G(\mathbb{A})$. To make sense of periods on π, we assume for the moment that the contribution of π to the intersection cohomology of $\mathrm{Sht}^\mu_{G,K}$ is as predicted in (3-3). Restricting $\mathcal{P}^{G,\mu}_{H,\lambda}$ to the π-part we get

$$\pi^K\otimes(\otimes_{i=1}^r \mathrm{H}_c^1((X-N)\otimes\overline{k},\rho_\pi^{\mu_i})) \xrightarrow{\theta^*} \mathrm{H}_c^{2d_H(\lambda)}(\mathrm{Sht}^\lambda_{H,K_H}\otimes\overline{k}) \to \overline{\mathbb{Q}}_\ell.$$

As we expect $\rho_\pi^{\mu_i}$ to be pure, the above map should factor through the pure quotient of $\mathrm{H}_c^1((X-N)\otimes\overline{k},\rho_\pi^{\mu_i})$, which is $\mathrm{H}_c^1(X\otimes\overline{k}, j_{!*}\rho_\pi^{\mu_i})$ (the cohomology of the middle extension of $\rho_\pi^{\mu_i}$), and which does not change after enlarging N. Now shrinking K and passing to the direct limit, we get

$$\mathcal{P}^{G,\mu}_{H,\lambda,\pi} : \pi\otimes(\otimes_{i=1}^r \mathrm{H}^1(X\otimes\overline{k}, j_{!*}\rho_\pi^{\mu_i})) \to \overline{\mathbb{Q}}_\ell$$

which is the analogue of the classical period (2-12). Again $\mathcal{P}^{G,\mu}_{H,\lambda,\pi}$ should factor through the coinvariants $\pi_{H(\mathbb{A})}\otimes(\cdots)$.

3.3 Shtuka version of relative trace formula.

3.3.1 The setup. Let H_1 and H_2 be reductive subgroups of G over k. Fix an integer $r\geq 0$. Let $\lambda=(\lambda_1,\cdots,\lambda_r)$ (resp. κ and μ) be an admissible sequence of dominant coweights of H_1 (resp. H_2 and G). Assume that $\lambda_i \leq_G \mu_i$ and $\kappa_i \leq_G \mu_i$. In this case there are natural morphisms

$$\mathrm{Sht}^{\lambda}_{H_1} \xrightarrow{\theta_1} \mathrm{Sht}^{\mu}_{G} \xleftarrow{\theta_2} \mathrm{Sht}^{\kappa}_{H_2}$$

Suppose

$$\dim \mathrm{Sht}^{\lambda}_{H_1} = \dim \mathrm{Sht}^{\kappa}_{H_2} = \frac{1}{2} \dim \mathrm{Sht}^{\mu}_{G}.$$

With the same extra assumptions as in Section 3.2.3, we may define the periods $\mathcal{P}^{G,\mu}_{H_1,\lambda,\pi}$ and $\mathcal{P}^{G,\mu}_{H_2,\kappa,\widetilde{\pi}}$ (where $\widetilde{\pi}$ is the contragradient of π). We expect the tensor product

$$\mathcal{P}^{G,\mu}_{H_1,\lambda,\pi} \otimes \mathcal{P}^{G,\mu}_{H_2,\kappa,\widetilde{\pi}} : \pi \otimes \widetilde{\pi} \otimes \Big(\otimes_{i=1}^{r} \Big(\mathrm{H}^1(X \otimes \overline{k}, j_{!*}\rho^{\mu_i}_{\pi}) \otimes \mathrm{H}^1(X \otimes \overline{k}, j_{!*}\rho^{\mu_i}_{\widetilde{\pi}})\Big)\Big) \to \overline{\mathbb{Q}}_{\ell}$$

to factor through the pairing between $\mathrm{H}^1(X \otimes \overline{k}, j_{!*}\rho^{\mu_i}_{\pi})$ and $\mathrm{H}^1(X \otimes \overline{k}, j_{!*}\rho^{\mu_i}_{\widetilde{\pi}})$ given by the cup product (the local systems $\rho^{\mu_i}_{\widetilde{\pi}}$ and $\rho^{\mu_i}_{\pi}$ are dual to each other up to a Tate twist). Assuming this, we get a pairing

$$\mathcal{P}^{G,\mu}_{H_1,\lambda,\pi} \otimes \mathcal{P}^{G,\mu}_{H_2,\kappa,\widetilde{\pi}} : \pi_{H_1(\mathbb{A})} \otimes \widetilde{\pi}_{H_2(\mathbb{A})} \to \overline{\mathbb{Q}}_{\ell}.$$

When $\dim \pi_{H_1(\mathbb{A})} = \dim \widetilde{\pi}_{H_2(\mathbb{A})} = 1$, we expect the ratio between the above pairing and the Petersson inner product to be related to *derivatives* of L-functions of π, though I do not know how to formulate a precise conjecture in general.

3.3.2 The relative trace. One way to access $\mathcal{P}^{G,\mu}_{H_1,\lambda,\pi} \otimes \mathcal{P}^{G,\mu}_{H_2,\kappa,\widetilde{\pi}}$ is to develop a relative trace formula whose spectral expansion gives these periods. Fix a compact open subgroup $K \subset G(\mathbb{A})$ and let $K_i = H_i(\mathbb{A}) \cap K$. Assume θ_i are proper, we have Heegner-Drinfeld cycles

$$\mathcal{Z}^{\lambda}_{H_1} = \theta_{1*}[\mathrm{Sht}^{\lambda}_{H_1,K_1}], \quad \mathcal{Z}^{\kappa}_{H_2} = \theta_{2*}[\mathrm{Sht}^{\kappa}_{H_2,K_2}]$$

both of half dimension in $\mathrm{Sht}^{\mu}_{G,K}$. Intuitively we would like to form the intersection number

(3-6) $$I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}(f) = \langle \mathcal{Z}^{\lambda}_{H_1}, f * \mathcal{Z}^{\kappa}_{H_2} \rangle_{\mathrm{Sht}^{\mu}_{G,K}}, \quad f \in C_c(K \backslash G(\mathbb{A})/K)$$

as the "relative trace" of f in this context. Here $f * (-)$ denotes the action of the Hecke algebra on the Chow group of $\mathrm{Sht}^{\mu}_{G,K}$, defined similarly as in Section 3.1.2. However, there are several technical issues before we can make sense of this intersection number.

1. $\mathrm{Sht}^{\mu}_{G,K}$ may not be smooth so the intersection product of cycles may not be defined.

2. Suppose the intersection of $\mathcal{Z}^{\lambda}_{H_1}$ and $\mathcal{Z}^{\kappa}_{H_2}$ is defined as a 0-cycle on $\mathrm{Sht}^{\mu}_{G,K}$, if we want to get a number out of this 0-cycle, we need it to be a proper cycle, i.e., it should lie in the Chow group of cycles with proper support (over k).

The first issue goes away if we assume each μ_i to be a minuscule coweight of G, which guarantees that $\mathrm{Sht}^{\mu}_{G,K}$ is smooth over $(X-N)^r$. The second issue is more serious and is analogous to the divergence issue for the usual relative trace. In results that we will present later, it won't be an issue because there $\mathcal{Z}^{\lambda}_{H_1,K_1}$ is itself a proper cycle. In the sequel we will proceed with heuristic arguments as we did in Section 2.2, and ignore these issues.

When λ, κ, μ are all zero, the linear functional $I^{(G,0)}_{(H_1,0),(H_2,0)}$ becomes the relative trace $\mathrm{RTr}^{G}_{H_1,H_2}$ defined in (2-14). Therefore the functional $I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}$ is a generalization of the relative trace.

3.3.3 Intersection number in terms of Hitchin-like moduli stacks. In the case of the usual relative trace for $f = \mathbf{1}_{KgK}$, we introduced a Hitchin-like moduli stack $\mathfrak{M}^{G}_{H_1,H_2,KgK}$ whose point-counting is essentially the relative trace of f. We now try to do the same for $I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}$. To simplify notations we assume $K = K_0 = \prod_x G(\mathcal{O}_x)$, hence $K_i = \prod H_i(\mathcal{O}_x)$, and suppress them from the notation for Shtukas.

To calculate the intersection number (3-6), a natural starting point is to form the stack-theoretic intersection of the cycles $\mathcal{Z}^{\lambda}_{H_1}$ and $f * \mathcal{Z}^{\kappa}_{H_2}$, i.e., consider the Cartesian diagram

$$\begin{array}{ccc} \mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0} & \longrightarrow & \mathrm{Sht}^{\mu}_{G,K_0gK_0} \\ \downarrow & & \downarrow \\ \mathrm{Sht}^{\lambda}_{H_1} \times \mathrm{Sht}^{\kappa}_{H_2} & \xrightarrow{\theta_1\times\theta_2} & \mathrm{Sht}^{\mu}_{G} \times \mathrm{Sht}^{\mu}_{G} \end{array} \tag{3-7}$$

The expected dimension of $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$ is zero. If $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$ indeed was zero-dimensional and moreover was proper over k, then $I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}(\mathbf{1}_{K_0gK_0})$ would be equal to the length of $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$. However, neither the zero-dimensionality nor the properness is true in general. Putting these issues aside, we proceed to rewrite $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$ in Hitchin-like terms.

Recall we have Hecke correspondences Hk^{μ}_{G}, $\mathrm{Hk}^{\lambda}_{H_1}$ and $\mathrm{Hk}^{\kappa}_{H_2}$ for Bun_G, Bun_{H_1} and Bun_{H_2} related by the maps $\theta_{1,\mathrm{Hk}}$ and $\theta_{2,\mathrm{Hk}}$. We can also define a Hecke correspondence for Hk_{G,K_0gK_0} as the moduli stack classifying $x_1, \cdots, x_r \in X$ and a commutative diagram of rational isomorphisms of G-bundles over X

(3-8)
$$\begin{array}{ccccccc} \mathcal{E}_0 & \dashrightarrow & \mathcal{E}_1 & \dashrightarrow & \cdots & \dashrightarrow & \mathcal{E}_r \\ \downarrow{\scriptstyle \varphi_0} & & \downarrow{\scriptstyle \varphi_1} & & & & \downarrow{\scriptstyle \varphi_r} \\ \mathcal{E}'_0 & \dashrightarrow & \mathcal{E}'_1 & \dashrightarrow & \cdots & \dashrightarrow & \mathcal{E}'_r \end{array}$$

such that

1. The top and bottom rows of the diagram give objects in Hk^{μ}_{G} over $(x_1, \cdots, x_r) \in X^r$;
2. Each column of the diagram gives an object in Hk_{G,K_0gK_0}, i.e., the relative position of φ_i is given by K_0gK_0 for $0 \leq i \leq r$.

We denote the resulting moduli stack by $\mathrm{Hk}^{\mu}_{\mathrm{Hk},K_0gK_0}$. We have maps $\mathfrak{p}, \mathfrak{q} : \mathrm{Hk}^{\mu}_{\mathrm{Hk},K_0gK_0} \to \mathrm{Hk}^{\mu}_{G}$ by taking the top and the bottom rows; we also have maps $\mathfrak{p}_i : \mathrm{Hk}^{\mu}_{\mathrm{Hk},K_0gK_0} \to \mathrm{Hk}_{G,K_0gK_0}$ by taking the ith column. The Hecke correspondence $\mathrm{Sht}^{\mu}_{G,K_0gK_0}$ in Section 3.1.2 is defined as the pullback of $(\mathfrak{p}_0, \mathfrak{p}_r) : \mathrm{Hk}^{\mu}_{\mathrm{Hk},K_0gK_0} \to \mathrm{Hk}_{G,K_0gK_0} \times \mathrm{Hk}_{G,K_0gK_0}$ along the graph of the Frobenius morphism for Hk_{G,K_0gK_0}.

We then define $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$ using the Cartesian diagram

$$\begin{array}{ccc} \mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0} & \longrightarrow & \mathrm{Hk}^{\mu}_{\mathrm{Hk},K_0gK_0} \\ \downarrow & & \downarrow{\scriptstyle (\mathfrak{p},\mathfrak{q})} \\ \mathrm{Hk}^{\lambda}_{H_1} \times \mathrm{Hk}^{\kappa}_{H_2} & \xrightarrow{\theta_{1,\mathrm{Hk}} \times \theta_{2,Hk}} & \mathrm{Hk}^{\mu}_{G} \times \mathrm{Hk}^{\mu}_{G} \end{array}$$

Now $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$ can be viewed as an r-step Hecke correspondence for $\mathfrak{M}^{G}_{H_1,H_2,K_0gK_0}$. Indeed, $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$ classifies a diagram similar to (3-8), except that $\mathcal{E}_i$ (resp. $\mathcal{E}'_i$) are now induced from H_1-bundles $\mathcal{F}_i$ (resp. H_2-bundles $\mathcal{F}'_i$), and the top row (resp. bottom row) are induced from an object in $\mathrm{Hk}^{\lambda}_{H_1}$ (resp. $\mathrm{Hk}^{\kappa}_{H_2}$). Now each column in such a diagram gives an object in $\mathfrak{M}^{G}_{H_1,H_2,K_0gK_0}$. Recording the i-th column gives a map m_i : $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0} \to \mathfrak{M}^{G}_{H_1,H_2,K_0gK_0}$. We claim that there is a Cartesian diagram expressing $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$ as the "moduli of Shtukas for $\mathfrak{M}^{G}_{H_1,H_2,K_0gK_0}$ with modification type (λ,κ)":

(3-9)
$$\begin{array}{ccc} \mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0} & \longrightarrow & \mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0} \\ \downarrow & & \downarrow{\scriptstyle (m_0,m_r)} \\ \mathfrak{M}^G_{H_1,H_2,K_0gK_0} & \xrightarrow{(\mathrm{id},\mathrm{Fr})} & \mathfrak{M}^G_{H_1,H_2,K_0gK_0}\times\mathfrak{M}^G_{H_1,H_2,K_0gK_0} \end{array}$$

Indeed, this diagram is obtained by unfolding each corner of (3-7) as a fiber product of an r-step Hecke correspondence with the graph of Frobenius, and re-arranging the order of taking fiber products.

Continuing with the heuristics, the intersection number $I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}(\mathbf{1}_{K_0gK_0})$, which "is" the length of $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$, should also be the intersection number of $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$ with the graph of Frobenius for $\mathfrak{M}^G_{H_1,H_2,K_0gK_0}$. In other words, we are changing the order of intersection and leaving the "Shtuka-like" construction to the very last step. It is often true that $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$ has the same dimension as $\mathfrak{M}^G_{H_1,H_2,K_0gK_0}$, and the fundamental class of $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$ induces an endomorphism of the cohomology of $\mathfrak{M}^G_{H_1,H_2,K_0gK_0}$ which we denote by $[\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}]$. The Lefschetz trace formula then gives the following heuristic identity

(3-10)
$$I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}(\mathbf{1}_{K_0gK_0})'' ='' \mathrm{Tr}([\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}]\circ\mathrm{Frob}, \mathrm{H}^*_c(\mathfrak{M}^G_{H_1,H_2,K_0gK_0}\otimes\overline{k},\mathbb{Q}_\ell)).$$

The equal sign above is in quotation marks for at least two reasons: both sides may diverge; change of the order of intersection needs to be justified.

Let $\mathfrak{B}^G_{H_1,H_2,K_0gK_0}$ be the Hitchin base and $h^G_{H_1,H_2}$ be the Hitchin map as in (2-18). Observe that for various $0\le i\le r$, the compositions $h^G_{H_1,H_2}\circ m_i : \mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}\to \mathfrak{B}^G_{H_1,H_2,K_0gK_0}$ are all the same. On the other hand, the Frobenius of $\mathfrak{M}^G_{H_1,H_2,K_0gK_0}$ covers the Frobenius of $\mathfrak{B}^G_{H_1,H_2,K_0gK_0}$. Therefore diagram (3-9) induces a map

$$\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}\to\mathfrak{B}^G_{H_1,H_2,K_0gK_0}(k)$$

which simply says that $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}$ decomposes into a disjoint union

(3-11)
$$\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}=\coprod_{a\in\mathfrak{B}^G_{H_1,H_2,K_0gK_0}(k)}\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa),K_0gK_0}(a).$$

The action of $[\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}]$ on the cohomology of $\mathfrak{M}^G_{H_1,H_2,K_0gK_0}$ can be localized to an action on the complex $\mathbf{R}h^G_{H_1,H_2,!}\mathbb{Q}_\ell$ using the formalism of cohomological correspondences. Then we may rewrite (3-10) as

$$I^{(G,\mu)}_{(H_1,\lambda),(H_2,\kappa)}(\mathbf{1}_{K_0gK_0})'' ='' \sum_a \mathrm{Tr}([\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}]_a \circ \mathrm{Frob}_a, (\mathbf{R}h^G_{H_1,H_2,!}\mathbb{Q}_\ell)_a), \tag{3-12}$$

where a runs over $\mathfrak{B}^G_{H_1,H_2,K_0gK_0}(k)$.

Comparing (3-12) with (2-20), we see the only difference is the insertion of the operator $[\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}]_a$ acting on the stalk $(\mathbf{R}h^G_{H_1,H_2,!}\mathbb{Q}_\ell)_a$.

3.3.4 Example. Suppose $G = \mathrm{PGL}_2$ and $H_1 = H_2 = T$ as in Example 3.2.2. Let $r \geq 0$ be even. We pick $\lambda, \kappa \in \{\pm 1\}^r$ with total sum zero, and form Sht^λ_T and Sht^κ_T. Let $\mu = (\mu_1, \cdots, \mu_r)$ consist of minuscule coweights of G. In this situation we have $\dim \mathrm{Sht}^\lambda_T = \dim \mathrm{Sht}^\kappa_T = r = \frac{1}{2}\dim \mathrm{Sht}^\mu_G$. Below we give more explicit descriptions of $\mathfrak{M}^G_{T,T,K_0gK_0}$ and $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},K_0gK_0}$. For simplicity, we assume that the double cover $\nu : X' \to X$ is étale.

As in Example 2.2.6, it is more convenient to work with the test function h_D rather than $\mathbf{1}_{K_0gK_0}$, where D is an effective divisor on X. We denote the corresponding version of $\mathfrak{M}^G_{T,T,K_0gK_0}$ by $\mathfrak{M}^G_{T,T,D}$. To describe $\mathfrak{M}^G_{T,T,D}$, we first consider the moduli stack $\widetilde{\mathfrak{M}}^G_{T,T,D}$ classifying $(\mathcal{L}, \mathcal{L}', \varphi)$ where $\mathcal{L}$ and $\mathcal{L}'$ are line bundles over X', and $\varphi : \nu_*\mathcal{L} \to \nu_*\mathcal{L}'$ is an injective map of coherent sheaves such that $\det(\varphi)$ has divisor D. We then have $\mathfrak{M}^G_{T,T,D} = \widetilde{\mathfrak{M}}^G_{T,T,D}/\mathrm{Pic}_X$, where Pic_X acts by pulling back to X' and simultaneously tensoring with $\mathcal{L}$ and $\mathcal{L}'$.

It is more convenient to work with another description of $\mathfrak{M}^G_{T,T,D}$. The map $\varphi : \nu_*\mathcal{L} \to \nu_*\mathcal{L}'$ is equivalent to the data of two maps

$$\alpha : \mathcal{L} \to \mathcal{L}', \quad \beta : \sigma^*\mathcal{L} \to \mathcal{L}'$$

where σ is the nontrivial involution of X' over X. The determinant $\det(\varphi) = \mathrm{Nm}(\alpha) - \mathrm{Nm}(\beta)$, as sections of $\mathrm{Nm}(\mathcal{L})^{-1} \otimes \mathrm{Nm}(\mathcal{L}')$. Let $\mathfrak{M}^{G,\diamond}_{T,T,D} \subset \mathfrak{M}^G_{T,T,D}$ be the open subset where α and β are nonzero. By recording the divisors of α and β, we may alternatively describe $\mathfrak{M}^{G,\diamond}_{T,T,D}$ as the moduli of pairs (D_α, D_β) of effective divisors on X' of degree $d = \deg D$, such that there exists a rational function f on X (necessarily unique) satisfying $\mathrm{div}(f) = \nu(D_\alpha) - \nu(D_\beta)$ and $\mathrm{div}(1-f) = D - \nu(D_\beta)$.

The Hecke correspondence $\mathrm{Hk}^{\lambda,\kappa}_{\mathfrak{M},D}$ is the composition of r correspondences each of which is either $\mathcal{H}_+$ or $\mathcal{H}_-$ depending on whether $\lambda_i = \kappa_i$ or not. Over the open subset $\mathfrak{M}^{G,\diamond}_{T,T,D}$, $\mathcal{H}_+$ can be described as follows: it classifies triples of effective divisors $(D_\alpha, D_\beta, D'_\beta)$ on X' such that $(D_\alpha, D_\beta) \in \mathfrak{M}^{G,\diamond}_{T,T,D}$, and D'_β is obtained by changing one point of D_β by its image under σ. The two maps $p_+, q_+ : \mathcal{H}_+ \to \mathfrak{M}^{G,\diamond}_{T,T,D}$ send $(D_\alpha, D_\beta, D'_\beta)$ to (D_α, D_β) and (D_α, D'_β). Similarly, over $\mathfrak{M}^{G,\diamond}_{T,T,D}$, $\mathcal{H}_-$ classifies triples of effective divisors $(D_\alpha, D'_\alpha, D_\beta)$ on X' such that D'_α is obtained by changing one point of D_α by its image under σ.

3.4 Application to L-functions. In the work Yun and Zhang [2017], we considered the case $G = \mathrm{PGL}_2$ and the moduli of Shtukas Sht^r_G without level structures, where r stands for the r-tuple $\mu = (\mu_1, \cdots, \mu_r)$ consisting of minuscule coweights of G (so r is even). Let $\nu : X' \to X$ be an unramified double cover. The Heegner-Drinfeld cycle we considered was the one introduced in Example 3.2.2, i.e., Sht^λ_T for $\lambda \in \{\pm 1\}^r$. We consider the lifting of the natural map $\theta : \mathrm{Sht}^\lambda_T \to \mathrm{Sht}^r_G$

$$\theta' : \mathrm{Sht}^\lambda_T \to \mathrm{Sht}'^r_G := \mathrm{Sht}^r_G \times_{X^r} X'^r.$$

Since Sht^λ_T is proper of dimension r, the Heegner-Drinfeld cycle $\mathcal{Z}^\lambda_T := \theta'_*[\mathrm{Sht}^\lambda_T]$ is an r-dimensional proper cycle in the $2r$-dimensional Sht'^r_G. Therefore $\mathcal{Z}^\lambda_T$ defines a class $Z^\lambda_T \in \mathrm{H}^{2r}_c(\mathrm{Sht}'^r_G \otimes \overline{k}, \overline{\mathbb{Q}}_\ell)(r)$.

Now let π be an everywhere unramified cuspidal automorphic representation of $G(\mathbb{A})$ with coefficients in $\overline{\mathbb{Q}}_\ell$. By the coarse cohomological spectral decomposition for $\mathrm{H}^{2r}_c(\mathrm{Sht}'^r_G \otimes \overline{k}, \overline{\mathbb{Q}}$ (see Theorem 3.1.5), we may project Z^λ_T to the χ_π-isotypical summand, and denote the resulting class by $Z^\lambda_{T,\pi} \in \mathrm{H}^{2r}_c(\mathrm{Sht}'^r_G \otimes \overline{k}, \overline{\mathbb{Q}}_\ell)[\chi_\pi]$.

3.4.1 Theorem (Yun and Zhang [ibid.]). *We have*

$$\langle Z^\lambda_{T,\pi}, Z^\lambda_{T,\pi}\rangle_{\mathrm{Sht}'^r_G} = \frac{q^{2-2g}}{2(\log q)^r}\frac{\mathcal{L}^{(r)}(\pi_{F'}, 1/2)}{L(\pi, \mathrm{Ad}, 1)}$$

where

- $\langle Z^\lambda_{T,\pi}, Z^\lambda_{T,\pi}\rangle_{\mathrm{Sht}'^r_G}$ *is the self-intersection number of the cycle class* $Z^\lambda_{T,\pi}$.
- $\pi_{F'}$ *is the base change of* π *to* $F' = k(X')$.
- $\mathcal{L}(\pi_{F'}, s) = q^{4(g-1)(s-1/2)} L(\pi_{F'}, s)$ *is the normalized* L*-function of* $\pi_{F'}$ *such that* $\mathcal{L}(\pi_{F'}, s) = \mathcal{L}(\pi_{F'}, 1-s)$.

In Yun and Zhang [n.d.], we extended the above theorem to allow the automorphic representation π to have square-free level structures (which means the local representations π_v are either unramified or an unramified twist of the Steinberg representation), and to allow ramifications for the double cover $\nu : X' \to X$. We consider the moduli stack $\mathrm{Sht}^r_G(\Sigma; \Sigma_\infty)$ where Σ is a finite set of places where we add Iwahori level structures to the G-Shtukas; $\Sigma_\infty \subset \Sigma$ is a subset of places where we impose supersingular conditions. The admissibility condition forces r to have the same parity as $\#\Sigma_\infty$.

3.4.2 Theorem (Yun and Zhang [ibid.]). *Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ with square-free level Σ. Assume the double cover $\nu : X' \to X$ is unramified over Σ. Let $\Sigma_\infty \subset \Sigma$ be the places that are inert in F'. Let $r \in \mathbb{Z}_{\geq 0}$ be of the same parity as $\#\Sigma_\infty$. Then for any $r_1, r_2 \in \mathbb{Z}_{\geq 0}$ such that $r_1 + r_2 = r$, there is an explicit linear combination $Z_T^{r_1,r_2}$ of the cycles $\{Z_T^\lambda; \lambda \in \{\pm 1\}^r\}$ such that*

$$\langle Z_{T,\pi}^{r_1,r_2}, Z_{T,\pi}^{r_1,r_2} \rangle_{\mathrm{Sht}_G'^r(\Sigma;\Sigma_\infty)} = \frac{q^{2-2g+\rho/2-N}}{2(-\log q)^r} \frac{\mathcal{L}^{(r_1)}(\pi, 1/2)\mathcal{L}^{(r_2)}(\pi \otimes \eta_{F'/F}, 1/2)}{L(\pi, \mathrm{Ad}, 1)}.$$

where

- $N = \deg \Sigma$*, and ρ is the degree of the ramification locus of ν.*
- $\mathcal{L}(\pi, s) = q^{(2g-2+N/2)(s-1/2)} L(\pi, s)$ *is the normalized L-function of π such that* $\mathcal{L}(\pi, s) = \mathcal{L}(\pi, 1-s)$.
- $\eta_{F'/F}$ *is the character of $F^\times \backslash \mathbb{A}^\times$ corresponding to the quadratic extension F'/F.*
- $\mathcal{L}(\pi \otimes \eta_{F'/F}, s) = q^{(2g-2+\rho+N/2)(s-1/2)} L(\pi \otimes \eta_{F'/F}, s)$ *is the normalized L-function of $\pi \otimes \eta_{F'/F}$ such that $\mathcal{L}(\pi \otimes \eta_{F'/F}, s) = \mathcal{L}(\pi \otimes \eta_{F'/F}, 1-s)$.*

When $r = 0$, the above theorem is a special case of the Waldspurger formula Waldspurger [1985], and our proof in this case is very close to the one in Jacquet [1986]. When $r = 1$ and $\#\Sigma_\infty = 1$, the above theorem is an analogue of the Gross-Zagier formula (see Gross and Zagier [1986]) which expresses the first derivative of the base-change L-function of a cuspidal Hecke eigenform in terms of the height of Heegner points on the modular curve. However our proof is very different from the original proof of the Gross-Zagier formula in that we do not need to explicitly compute either side of the formula.

3.4.3 Relation with the B-SD conjecture. Theorem 3.4.2 is applicable to those π coming from semistable elliptic curves E over the function field F. The relation of our result and the Birch–Swinnerton-Dyer conjecture for E can be roughly stated as follows. Take r to be the vanishing order of $L(E_{F'}, s) = L(\pi_{F'}, s-1/2)$ at $s = 1$. According to the expectation (3-3), $Z_{T,\pi}^\lambda$ is an element in $\pi^K \otimes \mathrm{H}^1(X' \otimes \overline{k}, j_{!*}\nu^*\rho_\pi)^{\otimes r}$. The 2-dimensional ℓ-adic Galois representation ρ_π attached to π is the Tate module of E, therefore $L(E_{F'}, s) = \det(1 - q^{-s}\mathrm{Frob}|\mathrm{H}^1(X' \otimes \overline{k}, j_{!*}\nu^*\rho_\pi))$. The standard conjecture predicts that the Frobenius acts semisimply on $\mathrm{H}^1(X' \otimes \overline{k}, j_{!*}\nu^*\rho_\pi)$, hence the multiplicity of the Frobenius eigenvalue q should be r. We expect $Z_{T,\pi}^\lambda$ to lie in $\pi^K \otimes \wedge^r(\mathrm{H}^1(X' \otimes \overline{k}, j_{!*}\nu^*\rho_\pi)^{\mathrm{Fr}=q})$, and giving a basis for this hypothetically 1-dimensional space. However, currently we do not have a way to construct rational points on E from the Heegner-Drinfeld cycle Sht_T^λ.

3.4.4 . The method to prove Theorems 3.4.1 is by comparing the Shtuka version of the relative trace $I^G_{T,T}(f) = \langle \mathcal{Z}^\lambda_T, f * \mathcal{Z}^\lambda_T \rangle_{\mathrm{Sht}'^r_G}$ as in (3-6) with the usual relative trace of the kind in Example 2.2.6. More precisely, for the triple (G, A, A) considered in Example 2.2.6, we consider the relative trace involving a complex variable s

$$\mathrm{RTr}^G_{A,(A,\eta)}(f,s) = \langle \varphi_! |\cdot|^s_{A(F)\backslash A(\mathbb{A})}, \varphi_!(\eta|\cdot|^s)_{A(F)\backslash A(\mathbb{A})} \rangle_{L^2(G(F)\backslash G(\mathbb{A}),\mu_G)}.$$

Here $|\cdot| : A(F)\backslash A(\mathbb{A}) = F^\times\backslash\mathbb{A}^\times \to q^{\mathbb{Z}}$ is the global absolute value function, and $\eta = \eta_{F'/F}$. Let $J_r(f)$ be the rth derivative of $\mathrm{RTr}^G_{A,(A,\eta)}(f,s)$ at $s = 0$. The key to the proof is to establish the following identity of relative traces for all spherical Hecke functions f

$$I^G_{T,T}(f) = (\log q)^{-r} J_r(f). \tag{3-13}$$

To prove this identity, it suffices to consider $f = h_D$ for effective divisors D on X (see Example 2.2.6). The moduli stacks $\mathfrak{M}^G_{A,A,D}$ in Example 2.2.6 and $\mathfrak{M}^G_{T,T,D}$ in Example 3.3.4 share the same Hitchin base $\mathcal{B}_D = \mathrm{H}^0(X, \mathcal{O}_X(D))$. We may fix a degree d and let D vary over effective divisors of degree d and get Hitchin maps $f_d : \mathfrak{M}^G_{A,A,d} \to \mathcal{B}_d$ and $g_d : \mathfrak{M}^G_{T,T,d} \to \mathcal{B}_d$. Formulae (2-20) and (3-12) suggest that we should try to prove an identity between the direct image complexes of f_d and g_d. The new geometric input here is the action of the Hecke correspondences $[\mathrm{Hk}^{\lambda,\lambda}_{\mathfrak{M}_d}]$ on the complex $\mathbf{R}g_{d!}\mathbb{Q}_\ell$, which is the r-th iteration of the action of the correspondence $[\mathcal{H}_+]$ defined in Example 3.3.4. It turns out that the eigenvalues of the action of $[\mathcal{H}_+]$ on $\mathbf{R}g_{d!}\mathbb{Q}_\ell$ match exactly with the factors coming from taking the derivative of the relative trace $\mathrm{RTr}^G_{A,(A,\eta)}(f,s)$, which explains why derivatives of automorphic quantities are indeed geometric.

3.5 Arithmetic fundamental lemma. Generalizing Theorems 3.4.1 and 3.4.2 to higher rank groups would involve intersecting non-proper cycles in an ambient stack which is not of finite type. This is the same issue as the non-convergence of the naive relative trace (2-14), therefore a certain truncation and regularization procedure is needed. There is, however, a local version of such results that can be proved for higher rank groups. One example of such a local version is the Arithmetic Fundamental Lemma formulated by W. Zhang [2012] originally for Rapoport-Zink spaces. In Yun [n.d.], we stated a higher derivative extension of W.Zhang's conjecture in the function field case, and sketched a proof. This was the first time higher derivatives of automorphic quantities were related to geometry, and it partially motivated the later work Yun and Zhang [2017].

3.5.1 Local Shtukas. The moduli of Shtukas has a local version. Fix a local function field F_x with ring of integers $\mathcal{O}_x$. In the diagram (3-2) defining the moduli of Shtukas, we may replace Bun_G by the affine Grassmannian Gr_G, and replace Hk_G^μ by an iterated Hecke correspondence for Gr_G over a formal disk $\Delta_r = \mathrm{Spf}(\mathcal{O}_x \widehat{\otimes} \cdots \widehat{\otimes} \mathcal{O}_x)$ of dimension r. We also have the freedom of changing the Frobenius morphism on Gr_G by the composition $b \circ \mathrm{Fr} : \mathrm{Gr}_G \to \mathrm{Gr}_G$, where b is an element of the loop group of G giving the datum of a G-isocrystal. The resulting object ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ by forming the Cartesian square as (3-2) is called the moduli space of local Shtukas, and it is a formal scheme over Δ_r. The special fiber of ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ is an iterated version of the affine Deligne-Lusztig variety. The twisted centralizer G_b of b is an inner form of a Levi subgroup of G, and $G_b(F_x)$ acts on ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$.

For a subgroup $H \subset G$ with a sequence of coweights $\lambda = (\lambda_1, \cdots, \lambda_r)$ such that $\lambda_i \leq_G \mu_i$ for all i and an H-isocrystal b_H compatible with b, we have a morphism $\theta^{\mathrm{loc}} : {}^{b_H}\mathrm{Sht}_H^{\lambda,\mathrm{loc}} \to {}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ over Δ_r. This morphism is a closed embedding because $\mathrm{Gr}_H \to \mathrm{Gr}_G$ is. We define the local Heegner-Drinfeld cycle ${}^{b_H}\mathcal{Z}_H^{\lambda,\mathrm{loc}}$ as the image of θ^{loc}.

If we have two local Heegner-Drinfeld cycles ${}^{b_1}\mathcal{Z}_{H_1}^{\lambda,\mathrm{loc}}$ and ${}^{b_2}\mathcal{Z}_{H_2}^{\kappa,\mathrm{loc}}$ in ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$ with complementary dimensions, and if μ_i are minuscule and the reduced structure of their intersection is proper over k, we may ask for their intersection number in ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$. More generally, if $\delta \in G_b(F_x)$, we may consider the intersection number

$$I_\delta = \langle {}^{b_1}\mathcal{Z}_{H_1}^{\lambda,\mathrm{loc}}, \delta \cdot {}^{b_2}\mathcal{Z}_{H_2}^{\kappa,\mathrm{loc}} \rangle_{{}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}}$$

using the action of $G_b(F_x)$ on ${}^b\mathrm{Sht}_G^{\mu,\mathrm{loc}}$. When $\mu = 0$, this is the same as the local orbital integral $J^G_{H_1,H_2,\delta}(\mathbf{1}_{G(\mathcal{O}_x)})$ (see (2-15)) for the relative trace formula of the triple (G, H_1, H_2).

3.5.2 Example. Let F'_x/F_x be an unramified quadratic extension, with ring of integers $\mathcal{O}'_x$ and residue field k'_x. Fix a Hermitian vector space $W_{n,x}$ of dimension n over x, and let U_n be the unitary group of $W_{n,x}$. We define the moduli of local Shtukas $\mathrm{Sht}_{\mathrm{U}_n}^{r,\mathrm{loc}}$ over $\Delta'_r = \mathrm{Spf}(\mathcal{O}'_x \widehat{\otimes}_{k'_x} \cdots \widehat{\otimes}_{k'_x} \mathcal{O}'_x)$ in the following way. Let $\mathrm{Gr}_{\mathrm{U}_n}$ be the affine Grassmannian classifying self-dual lattices in $W_{n,x}$. Since U_n is split over F'_x, the base change $\mathrm{Gr}_{\mathrm{U}_n} \otimes_{k_x} k'_x$ can be identified with the affine Grassmannian $\mathrm{Gr}_{\mathrm{GL}_n} \otimes_{k_x} k'_x$ classifying $\mathcal{O}'_x$-lattices in $W_{n,x}$. We have the local Hecke correspondence $\mathrm{Hk}_{\mathrm{U}_n}^{\mathrm{loc}}$ over Δ'_1 which, after identifying $\mathrm{Gr}_{\mathrm{U}_n} \otimes_{k_x} k'_x$ with $\mathrm{Gr}_{\mathrm{GL}_n} \otimes_{k_x} k'_x$, corresponds to the upper modification of lattices in $W_{n,x}$ of colength one. Let $\mathrm{Hk}_{\mathrm{U}_n}^{r,\mathrm{loc}}$ be the r-fold composition of $\mathrm{Hk}_{\mathrm{U}_n}^{\mathrm{loc}}$ as a correspondence, so $\mathrm{Hk}_{\mathrm{U}_n}^{r,\mathrm{loc}} \to \Delta'_r$. Then $\mathrm{Sht}_{\mathrm{U}_n}^{r,\mathrm{loc}}$ is defined using the Cartesian diagram

$$\begin{array}{ccc} \mathrm{Sht}_{\mathrm{U}_n}^{r,\mathrm{loc}} & \longrightarrow & \mathrm{Hk}_{\mathrm{U}_n}^{r,\mathrm{loc}} \\ \downarrow & & \downarrow \\ \mathrm{Gr}_{\mathrm{U}_n} & \xrightarrow{(\mathrm{id},\mathrm{Fr})} & \mathrm{Gr}_{\mathrm{U}_n} \times \mathrm{Gr}_{\mathrm{U}_n} \end{array}$$

The group $\mathrm{U}_n(F_x)$ acts on $\mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_n}$. We remark that $\mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_n}$ is formally smooth over Δ'_r of relative dimension $r(n-1)$.

Now suppose we are in the situation of Example 2.3.4 so we have an orthogonal decomposition $W_{n,x} = W_{n-1,x} \oplus F'_x e_n$ and $(e_n, e_n) = 1$. Adding a standard lattice $\mathcal{O}'_x e_n$ gives an embedding $\mathrm{Gr}_{\mathrm{U}_{n-1}} \hookrightarrow \mathrm{Gr}_{\mathrm{U}_n}$ compatible with the Hecke modifications, hence induces an embedding $\mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_{n-1}} \hookrightarrow \mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_n}$. Consider the diagonal map

$$\Delta : \mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_{n-1}} \to \mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_{n-1}} \times_{\Delta'_r} \mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_n}.$$

The image of Δ gives an $r(n-1)$-dimensional cycle $\mathcal{Z}^{r,\mathrm{loc}}_x$ in the $2r(n-1)$-dimensional ambient space $\mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_{n-1}} \times_{\Delta'_r} \mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_n}$. For strongly regular semisimple $\delta \in \mathrm{U}_n(F_x)$ (with respect to the conjugation action by U_{n-1}) we form the intersection number

$$I^{\mathrm{U},r}_{x,\delta} = \langle \mathcal{Z}^{r,\mathrm{loc}}_x, (\mathrm{id} \times \delta)\mathcal{Z}^{r,\mathrm{loc}}_x \rangle_{\mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_{n-1}} \times_{\Delta'_r} \mathrm{Sht}^{r,\mathrm{loc}}_{\mathrm{U}_n}}.$$

This makes sense because the support of the intersection of the two cycles $\mathcal{Z}^{r,\mathrm{loc}}_x$ and $(\mathrm{id} \times \delta)\mathcal{Z}^{r,\mathrm{loc}}_x$ is proper. When $r = 0$, we have $I^{\mathrm{U},r}_{x,\delta}$ is either equal to $J^{\mathrm{U}}_{x,\delta}(\mathbf{1}_{\mathrm{U}(\Lambda_{n,x})})$ as in Example 2.3.4 if a self-dual lattice $\Lambda_{n,x} \subset W_{n,x}$ exists, or 0 otherwise.

To state the higher arithmetic fundamental lemma, we introduce a variant of the orbital integral (2-21) with a complex variable s

$$J^{\mathrm{GL}}_{x,\gamma}(f, s) = \int_{\mathrm{GL}_{n-1}(F_x)} f(h^{-1}\gamma h)\eta_x(\det h)|\det h|^s dh, \; \gamma \in \mathrm{S}_n(F_x), f \in C^\infty_c(\mathrm{S}_n(F_x)).$$

For γ strongly regular semisimple, $J^{\mathrm{GL}}_{x,\gamma}(f, s)$ is a Laurent polynomial in q_x^s, where $q_x = \#k_x$. Let

$$J^{\mathrm{GL},r}_{x,\gamma}(f) = \left(\frac{d}{ds}\right)^r |_{s=0} J^{\mathrm{GL}}_{x,\gamma}(f, s).$$

3.5.3 Theorem (Yun [n.d.]). *Let $\gamma \in \mathrm{GL}_n(F_x)$ and $\delta \in \mathrm{U}_n(F_x)$ be strongly regular semisimple with the same invariants in the sense of Example 2.3.3 and Example 2.3.4. Then*

$$I^{\mathrm{U},r}_{x,\delta} = c(\log q_x)^{-r} J^{\mathrm{GL},r}_{x,\gamma}(\mathbf{1}_{\mathrm{S}_n(\mathcal{O}_x)}) \tag{3-14}$$

with an explicit constant c depending on r and the invariants of γ.

The proof consists of the following main steps.

1. Prove a global analogue of (3-14). Consider the triple (G, H_1, H_2) as in Example 2.3.4 with coweights μ for G and λ for $H_1 = H_2$ being the first fundamental coweights. Let the triple (G', H'_1, H'_2) be as in Example 2.3.3. Recall that for

the global situation, the intersection number $I^{(G,\mu)}_{(H_1,\lambda),(H_2,\lambda)}(\mathbf{1}_{K_0 g K_0})$ is the degree of a certain 0-cycle on $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\lambda),K_0 g K_0}$ introduced in the diagram (3-7). On the other hand, we have a decomposition (3-11) of $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\lambda),K_0 g K_0}$ into a disjoint union of $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\lambda),K_0 g K_0}(a)$ indexed by k-points of the base $\mathcal{B}^G_{H_1,H_2,K_0 g K_0}$. For strongly regular semisimple a, $\mathrm{Sht}^{(G,\mu)}_{(H_1,\lambda),(H_2,\lambda),K_0 g K_0}(a)$ is proper for any g, so we can talk about the degree of the a-component of the zero cycle $\mathcal{Z}^\lambda_{H_1}\cdot(f * \mathcal{Z}^\lambda_{H_2})$, denoted $\langle \mathcal{Z}^\lambda_{H_1}, f * \mathcal{Z}^\lambda_{H_2}\rangle_a$. The global analogue of (3-14) means proving an identity of the form (3-13), but with both sides replaced by their a-components. One can prove such a global identity by analyzing the direct image complexes of the Hitchin maps $h^G_{H_1,H_2}$ and $h^{G'}_{H'_1,H'_2}$ using sheaf-theoretic methods, as we did in Yun [2011b] and Yun and Zhang [2017].

2. Deduce the arithmetic fundamental lemma from the global identity. The moduli $\mathrm{Sht}^{\mu,\mathrm{loc}}_G$ of local Shtukas for G is related to a formal completion of the global moduli stack Sht^μ_G by a uniformization diagram, analogous to the one relating Rapoport-Zink spaces and Shimura varieties. Using the uniformization, one can express $\langle \mathcal{Z}^\lambda_{H_1}, f * \mathcal{Z}^\lambda_{H_2}\rangle_a$ as a finite sum, where each summand is a product of usual orbital integrals and intersection numbers of the form $I^{\mathrm{U},r_i}_{x_i,\delta}$ (with $\sum r_i = r$). There is a similar product formula for the global orbital integral for (G', H'_1, H'_2). By choosing a appropriately we may deduce the local identity (3-14) from the global one using the product expansions and the known fundamental lemma (Theorem 2.3.5).

References

James Arthur (2005). "An introduction to the trace formula". In: *Harmonic analysis, the trace formula, and Shimura varieties*. Vol. 4. Clay Math. Proc. Amer. Math. Soc., Providence, RI, pp. 1–263. MR: 2192011 (cit. on p. 1468).

A. Bouthier, B. C. Ngô, and Y. Sakellaridis (2016). "On the formal arc space of a reductive monoid". *Amer. J. Math.* 138.1, pp. 81–108. MR: 3462881 (cit. on p. 1466).

Alexis Bouthier (2017). "La fibration de Hitchin-Frenkel-Ngô et son complexe d'intersection". *Ann. Sci. Éc. Norm. Supér. (4)* 50.1, pp. 85–129. MR: 3621427 (cit. on p. 1466).

Alexander Braverman and Roman Bezrukavnikov (2007). "Geometric Langlands correspondence for $\mathcal{D}$-modules in prime characteristic: the $\mathrm{GL}(n)$ case". *Pure Appl. Math. Q.* 3.1, Special Issue: In honor of Robert D. MacPherson. Part 3, pp. 153–179. MR: 2330157 (cit. on p. 1466).

Tsao-Hsien Chen and Xinwen Zhu (2017). "Geometric Langlands in prime characteristic". *Compos. Math.* 153.2, pp. 395–452. MR: 3705230 (cit. on p. 1466).

V. G. Drinfeld (1974). "Elliptic modules". *Mat. Sb. (N.S.)* 94(136), pp. 594–627, 656. MR: 0384707 (cit. on p. 1479).

– (1987). "Moduli varieties of F-sheaves". *Funktsional. Anal. i Prilozhen.* 21.2, pp. 23–41. MR: 902291 (cit. on p. 1479).

Edward Frenkel and Bao Châu Ngô (2011). "Geometrization of trace formulas". *Bull. Math. Sci.* 1.1, pp. 129–199. MR: 2823791 (cit. on p. 1468).

Benedict H. Gross and Don B. Zagier (1986). "Heegner points and derivatives of L-series". *Invent. Math.* 84.2, pp. 225–320. MR: 833192 (cit. on p. 1491).

Nigel Hitchin (1987). "Stable bundles and integrable systems". *Duke Math. J.* 54.1, pp. 91–114. MR: 885778 (cit. on p. 1465).

Hervé Jacquet (1986). "Sur un résultat de Waldspurger". *Ann. Sci. École Norm. Sup. (4)* 19.2, pp. 185–229. MR: 868299 (cit. on p. 1491).

Hervé Jacquet and Stephen Rallis (2011). "On the Gross-Prasad conjecture for unitary groups". In: *On certain L-functions*. Vol. 13. Clay Math. Proc. Amer. Math. Soc., Providence, RI, pp. 205–264. MR: 2767518 (cit. on p. 1476).

Anton Kapustin and Edward Witten (2007). "Electric-magnetic duality and the geometric Langlands program". *Commun. Number Theory Phys.* 1.1, pp. 1–236. MR: 2306566 (cit. on p. 1466).

Laurent Lafforgue (2002). "Chtoucas de Drinfeld et correspondance de Langlands". *Invent. Math.* 147.1, pp. 1–241. MR: 1875184 (cit. on pp. 1479, 1481).

Vincent Lafforgue (Sept. 2012). "Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale". arXiv: 1209.5352 (cit. on pp. 1479, 1480).

Bao Châu Ngô (2006). "Fibration de Hitchin et endoscopie". *Inventiones Math.* 164.2, pp. 399–453. MR: 2218781 (cit. on pp. 1466, 1468).

– (2010). "Le lemme fondamental pour les algèbres de Lie". *Publ. Math. Inst. Hautes Études Sci.* 111, pp. 1–169. MR: 2653248 (cit. on pp. 1466, 1471).

– (2014). "On a certain sum of automorphic L-functions". In: *Automorphic forms and related geometry: assessing the legacy of I. I. Piatetski-Shapiro*. Vol. 614. Contemp. Math. Amer. Math. Soc., Providence, RI, pp. 337–343. MR: 3220933 (cit. on p. 1466).

Alexei Oblomkov and Zhiwei Yun (2016). "Geometric representations of graded and rational Cherednik algebras". *Adv. Math.* 292, pp. 601–706. MR: 3464031 (cit. on p. 1466).

Y. Sakellaridis and A. Venkatesh (n.d.). "Periods and harmonic analysis on spherical varieties". To appear in Astérisque (cit. on p. 1472).

Yakov Varshavsky (2004). "Moduli spaces of principal F-bundles". *Selecta Math. (N.S.)* 10.1, pp. 131–166. MR: 2061225 (cit. on p. 1479).

J.-L. Waldspurger (1985). "Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie". *Compositio Math.* 54.2, pp. 173–242. MR: 783511 (cit. on p. 1491).

Zhiwei Yun (n.d.). "The arithmetic fundamental lemma over function fields" (cit. on pp. 1467, 1479, 1492, 1494).

– (2011a). "Global Springer theory". *Adv. Math.* 228.1, pp. 266–328. MR: 2822234 (cit. on p. 1466).

– (2011b). "The fundamental lemma of Jacquet and Rallis". *Duke Math. J.* 156.2. With an appendix by Julia Gordon, pp. 167–227. MR: 2769216 (cit. on pp. 1466, 1476, 1478, 1495).

– (2012). "Langlands duality and global Springer theory". *Compositio Math.* 148.3, pp. 835–867. MR: 2925401 (cit. on p. 1466).

– (2014). "The spherical part of the local and global Springer actions". *Math. Ann.* 359.3-4, pp. 557–594. MR: 3231007.

Zhiwei Yun and Wei Zhang (n.d.). "Shtukas and the Taylor expansion of L-functions, II". arXiv:1712.08026 (cit. on pp. 1467, 1479, 1482, 1490, 1491).

– (2017). "Shtukas and the Taylor expansion of L-functions". *Ann. of Math. (2)* 186.3, pp. 767–911. MR: 3702678 (cit. on pp. 1467, 1479, 1482, 1490, 1492, 1495).

Wei Zhang (2012). "On arithmetic fundamental lemmas". *Invent. Math.* 188.1, pp. 197–252. MR: 2897697 (cit. on p. 1492).

– (2014). "Fourier transform and the global Gan-Gross-Prasad conjecture for unitary groups". *Ann. of Math. (2)* 180.3, pp. 971–1049. MR: 3245011 (cit. on p. 1476).

Received 2017-11-30.

Zhiwei Yun (恽之玮)
Department of Mathematics, MIT, 77 Massachusetts Ave, Cambridge, MA 02139
zyun@mit.edu

Index of Authors

World Scientific
www.worldscientific.com
11060 hc

ISBN 978-981-3272-87-3 (set)
ISBN 978-981-3272-91-0 (v2)